

03063  
14



UNIVERSIDAD NACIONAL  
AUTÓNOMA DE  
MÉXICO

**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

**“DISEÑO Y ENTRENAMIENTO DE CLASIFICADORES  
NEURONALES ARTIFICIALES PARA EL  
RECONOCIMIENTO DE PATRONES DE DESARROLLO  
CELULAR MEDIANTE UNA METODOLOGÍA EVOLUTIVA  
MULTIOBJETIVO”**

**T E S I S**

QUE PARA OBTENER EL GRADO DE:

**MAESTRO EN CIENCIAS  
(COMPUTACIÓN)**

**P R E S E N T A:**

**JOSE ALEJANDRO ROSENDO ROBLES**

DIRECTOR DE TESIS:

**M. EN C. ALBERTO ARTURO HERRERA BECERRA**

TESIS CON  
FALLA DE ORIGEN



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## DEDICATORIAS

A Lupita Robles González

Cuando estoy lejos de ti tus bendiciones y rezos van conmigo, me recuerdan la fragilidad de tu abrazo y me hacen sentir la ternura infinita de tu mirada.

Te quiero mucho Mami.

A Rubén Rosendo Juárez

Tus consejos siempre me acompañan y tu comprensión me da toda la tranquilidad que necesito. Gracias por respetar mis silencios. En mis momentos de felicidad siempre recuerdo tu sonrisa.

Gracias Papá

A mi hermana Susana

tu nobleza me ha demostrado que, aun significando un esfuerzo se debe ayudar a cualquier persona sin esperar retribución. Y que todas las personas son buenas hasta que demuestran lo contrario.

A mi hermano Guillermo

espero algún día ser tan dedicado y constante como tu.

A mi hermana Lourdes

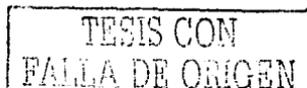
el ejemplo de tu sencillez me ha enseñado que la lealtad y el agradecimiento es la mejor forma de pagar un favor.

A mi carnal Juan

tu generosidad me ha mostrado cuanto se puede dar a los demás, incluso sacrificándose uno mismo.

A mi hermana Patricia

tu seriedad me ha enseñado que el respeto y la cortesía deben ser utilizadas disciplinadamente con todas las personas.



## AGRADECIMIENTOS

Al M.C. Alberto Arturo Herrera Becerra por darme la oportunidad de desarrollarme en lo que me gusta hacer. Todas sus indicaciones son acertadas y sustanciosas. Espero haber correspondido a su confianza.

A mis compañeros en el CCADET Alejandro Padrón, Juan Carlos Cuevas, Luis Ochoa, Rafael Prieto, Xiangjie Liu, por su ejemplo de dedicación y disciplina.

A mis sinodales, gracias por contribuir tanto con sus conocimientos para enriquecer mi trabajo. Gracias por dedicarme con cortesía su tiempo. Sus indicaciones van más allá de lo académico.

A todos mis profesores en la Maestría.... es un honor.

A todos mis compañeros de la Maestría,... no únicamente los profesores dan lecciones.

Gracias a todo el personal de IIMAS y del CCADET sus atenciones me hicieron sentir siempre entre amigos.

A mis amigos Socorro Téllez y Juan José Alcocer, gracias por su comprensión, cariño apoyo y respeto. Así es más fácil estar aquí.

TESIS CON  
FALLA DE ORIGEN

C

## RESUMEN

Las etapas de desarrollo de una célula pueden ser identificadas mediante una imagen que puede ser transformada en una serie de valores numéricos. Tenemos así un patrón que caracteriza una etapa del desarrollo de una célula. Las muestras celulares experimentales presentan cientos, incluso miles de células en distintos estados de desarrollo celular. Saber la frecuencia de ciertas etapas de desarrollo celular que se presentan en una muestra celular experimental es de gran utilidad, sobre todo porque en la naturaleza los ciclos celulares tienen determinada una frecuencia. Cuando éstas frecuencias de desarrollo se perturban se pueden desencadenar procesos que resultan en alteraciones graves en los organismos.

Obtener de una muestra celular experimental las frecuencias de ciertos estados celulares se reduce a un problema de identificar y contar. Contar en poco tiempo e identificar con altos niveles de confianza las alteraciones en las frecuencias de las distintas etapas de desarrollo celular de una muestra celular experimental significa diagnosticar a tiempo y correctamente ciertos padecimientos.

La identificación de las diferentes etapas de desarrollo celular no es una tarea sencilla, las distintas etapas de desarrollo celulares son todas y cada una parte de un proceso integral. Los patrones que identifican las distintas etapas más importantes del desarrollo celular no son fácilmente distinguibles, incluso para el experto entrenado en el tratamiento de muestras celulares. Las Redes Neuronales son buenos mecanismos para las tareas de identificación, tienen la capacidad de extraer información no explícita de las relaciones de una muestra de patrones.

El experto en redes neuronales sabe que las neuronas artificiales, deben interactuar complejamente para que puedan efectuarse los procesos que hacen posible el funcionamiento de una red neuronal. Esta complicidad simultánea entre interacción y complejidad implica lidiar con un gran número de grados de libertad que generalmente inducen amplitud y complejidad en el espacio de búsqueda del diseño y ajuste de redes neuronales.

0

TESIS CON  
FALLA DE ORIGEN

En la calibración de una Red Neuronal Clasificadora nunca se puede hacer una sola cosa y el curso de acción más apropiado a seguir viene determinado por los acontecimientos posteriores sobre todo en las Redes Neuronales supervisadas como es nuestro caso.

El gui3n conductor de este trabajo es una propuesta una metodolog3a no convencional que automatiza y optimiza simult3neamente el dise1o y ajuste de clasificadores neuronales artificiales para el reconocimiento de patrones de etapas de desarrollo celular, tomando como base las t3cnicas evolutivas.

Establecemos modificaciones al Algoritmo Gen3tico Simple. Determinamos una propia estructura cromos3mica heterog3nea y en consecuencia proponemos nuevos operadores gen3ticos. Combinamos t3cnicas alternas, como el ordenamiento de Fonseca, multipoblaciones y procesos de migraci3n, con t3cnicas usuales en los algoritmos gen3ticos, porque la combinaci3n de t3cnicas nos ofrece ventajas en comparaci3n con las metodolog3as usuales. El principal elemento del elenco anterior que conforma nuestra propuesta metodol3gica es una funci3n multiobjetivo que se revelo durante el an3lisis de la matriz de confusi3n. La funci3n multiobjetivo que proponemos ofrece de forma simult3nea m3tricas cualitativas y cuantitativas para la evaluaci3n correcta de los clasificadores neuronales artificiales desarrollados.

TESIS CON  
FALLA DE ORIGEN

E'

## INDICE GENERAL

### CAPÍTULO

Página

#### INTRODUCCIÓN

Motivación	i
Marco de Estudio	iv
Antecedentes	viii
Objetivos	xi
Aportaciones	xiii
Contenido de la Tesis	xiii

### UNO PROBLEMÁTICA DEL DISEÑO Y ENTRENAMIENTO DE UN CLASIFICADOR NEURONAL ARTIFICIAL (CNA) PARA EL RECONOCIMIENTO DE PATRONES DE ETAPAS DE DESARROLLO CELULAR

1.1	Distribuciones Estadísticas de los Patrones de Desarrollo Celular	1
1.2	Sesgo Implícito en la Población Experimental de Patrones	10
1.3	Retropropagación	12
1.4	Computación Evolutiva	16
1.5	Optimización Multiobjetivo	18
1.6	Deriva Genética	22
1.7	Arquitectura Básica del CNA	24
1.8	Resumen	25

TESIS CON  
FALLA DE ORIGEN

F

## **DOS METODOLOGÍA INTEGRAL PARA EL DISEÑO Y ENTRENAMIENTO DE CNAs MEDIANTE UN ALGORITMO EVOLUTIVO MULTIOBLACIONAL Y MULTIOBJETIVO**

2.1	Metodología	27
2.2	Muestreo y Representatividad	28
2.3	Descripción del Algoritmo Evolutivo Multipoblacional	34
2.3.1	Descripción de la Estructura Cromosómica no Homogénea	36
2.3.2	Matriz de Confusión y Función Multiobjetivo	39
2.3.3	Multipoblación y Migración	42
2.3.4	Asignación del Valor de Aptitud Relativo	46
2.3.5	Asignación de la Probabilidad de Selección Proporcional al Valor de Aptitud Relativo	48
2.3.6	Operadores de Cruzamiento y Mutación	50
2.3.7	Reinserción	54
2.3.8	Resumen	55

## **TRES EXPERIMENTACIÓN Y ANÁLISIS**

3.1	Obtención de Resultados	57
3.2	Análisis	59
3.2.1	Desarrollo y Evolución con una Muestra Homogénea	61
3.2.2	Desarrollo y Evolución con una Muestra Heterogénea	90
3.3	Comparación con otros Métodos	100

TESIS CON  
FALLA DE ORIGEN

<b>CUATRO</b>	<b>CONCLUSIONES</b>	
4.1	Discusión	105
4.2	Muestreo del Conjunto de Entrenamiento y del Conjunto de Validación	105
4.3	Funciones de Activación	106
4.4	Sobre y Bajo Ajuste	107
4.5	Función Objetivo	108
4.6	Metodología	109

**REFERENCIAS**

113

H

TESIS CON  
FALLA DE ORIGEN

*"La inteligencia... es la facultad de hacer objetos artificiales, sobre todo herramientas para hacer herramientas"*

**Henry Bergson**

I

TESIS CON  
FALLA DE ORIGEN

## INTRODUCCIÓN

### 1.1 MOTIVACIÓN

El diseño es considerado una actividad inteligente, en la cual el ser humano, busca nuevas soluciones para producir soluciones creativas. Dependiendo de la disciplina humana, el diseño es una actividad en menor o mayor grado compleja, dado que usualmente las condiciones que determinan el problema a resolver no son explícitas, existe carencia, incertidumbre y ambigüedad de información inicial, las restricciones del problema no son definidas, se tienen múltiples criterios que deben ser cumplidos y existen múltiples soluciones posibles. En específico, determinar la arquitectura de una red neuronal para resolver un problema de características lineales o no lineales, depende de las características propias del problema, de la experiencia y habilidad adquiridas por el experto en redes neuronales durante su investigación de sistemas adaptables[SM193]. Dado que el diseñador de redes neuronales debe tomar en cuenta que los sistemas con los que trabaja son sensibles a las condiciones iniciales, que los mecanismos y elementos utilizados por la red neuronal para autoajustarse poseen características particulares, que el tamaño y configuración de la arquitectura de red neuronal tiene consecuencias en su capacidad para procesar información y además aunado a esto, el experto en redes neuronales está conciente que todas las condiciones anteriores generalmente entran en conflicto. Así entonces para un problema determinado, la búsqueda de la topología y configuración de red neuronal óptima, en el sentido que conjunte simplicidad y eficiencia, es un problema complejo y debe ser manejado mediante un enfoque multiobjetivo. Esta problemática puede ser abordada tomando en cuenta tres principales elementos:

- Definir un número finito de variables de diseño, que operen en un rango delimitado de valores.
- Establecer un mecanismo que cuente con una cierta autonomía y sea capaz de generar e improvisar nuevas soluciones.
- Proponer una estrategia de toma de decisiones que nos proporcione como resultado una solución que cumpla con todos los criterios de forma simultánea.

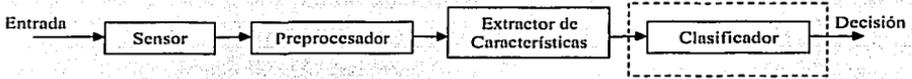


Figura 1.1 Componentes de un sistema identificador de patrones

La función principal de un sistema identificador de patrones es determinar correctamente las decisiones correspondientes a las clases de membresía de los patrones con los cuales es probado[GON78]. Muchos procesos importantes de manipulación y traslado de información, son llevados a cabo en el mismo tiempo en que un patrón entra al sistema y una decisión es tomada por el sistema. El proceso, resumido mediante diagrama de bloques, se ilustra en la figura 1.1.

El sensor es sencillamente el dispositivo de medición, que transforma el patrón de entrada a un formato adecuado para la manipulación del sistema. Algunos sistemas identificadores de patrones operan con la entrada directamente obtenida del sensor, aunque es práctica común complementar el sensor con un preprocesador y un extractor de características. El preprocesador, usualmente mediante alguna heurística que involucra algún conocimiento del dominio de los datos, u obedeciendo algunos criterios estadísticos y/o probabilísticos, remueve elementos innecesarios o corruptos de los datos medidos y posteriormente los normaliza adecuadamente en un intervalo determinado de valores. El extractor de características obtiene, de los datos preprocesados, las características necesarias para la clasificación. Finalmente esas características son la entrada para el clasificador, cuya función es determinar una decisión concerniente a la clase de membresía del patrón que está siendo procesado.

Un enfoque para la teoría de decisión, en reconocimiento de patrones, trata con algoritmos para estimar ciertos parámetros usando muestras de patrones en un proceso de entrenamiento. Si el sistema de reconocimiento de patrones se desempeña bien durante la etapa de entrenamiento, con la selección correcta de un conjunto representativo de patrones, la probabilidad de que el sistema, ya en operación, se desempeñe satisfactoriamente con datos de validación, incluso con datos directos de campo, se incrementa.

Dentro de la computación adaptable, existen numerosos algoritmos que son utilizados para obtener un sistema identificador de patrones. Esos algoritmos están divididos en:

- Determinantes, los cuales manejan la estimación de una función de decisión cuyos coeficientes están relacionados directamente con los patrones, sin haber recurrido a consideraciones estadísticas de los patrones (clasificadores basados en propiedades geométricas, clasificadores basados en la vecindad, clasificadores basados en lógica difusa, entre otros[[CHE99](#)], [[RAM99](#)], [[WEL94](#)]).
- Estadísticos, como su nombre lo indica, están basados en las propiedades estadísticas de la población de patrones para ajustar los parámetros de una función de decisión.

Dentro de los algoritmos estadísticos el ajuste de los parámetros de la función de decisión, es llevado a cabo específicamente utilizando muestras de patrones, lo que se propiamente se llama proceso de aprendizaje o entrenamiento. Las técnicas de aprendizaje de maquina pueden ser subdivididas en dos categorías principales:

- El aprendizaje supervisado, los parámetros del sistema clasificador son estimados mediante algoritmos que utilizan muestras de entrenamiento cuyas clases de las relaciones de membresía de los patrones están especificadas externamente por el diseñador del clasificador. De esta manera, los parámetros desconocidos del clasificador son ajustados para adaptarse a una situación en donde las clases de los patrones están especificadas y caracterizadas por muestras representativas de patrones. Es claro que el éxito final de este enfoque es dictado por la calidad de la muestra de entrenamiento empleada para entrenar el sistema de clasificación de patrones ([Adaline](#), [Retropropagación](#), entre otros[[SMI94](#)], [[WEL94](#)]).
- El enfoque de aprendizaje no supervisado es usado cuando no hay o es muy escaso un conocimiento a priori acerca de las clases de los patrones de un problema dado. En esencia este enfoque trata de extraer de los patrones, las clases presentes en un

conjunto de datos, para los cuales la clasificación de los patrones de la muestra disponible no está completamente conocida (las redes de Hamming, los mapas autoorganizativos, entre otros [LEO98], [CHI96]).

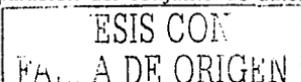
El objetivo principal de este trabajo es definir una metodología para diseñar y entrenar Clasificadores Neuronales Artificiales (CNAs) basados en aprendizaje supervisado, el resultado será una herramienta que auxilie al diseñador ofreciendo una variedad de CNAs que han sido diseñados y entrenados tomando en cuenta una serie de determinados criterios simultáneos.

La metodología resultante se probará en un caso particular: el diseño y entrenamiento de un clasificador neuronal de células en diferentes estados de desarrollo. El propósito del clasificador es determinar, sobre un conjunto de patrones digitales, cuáles pertenecen a determinados estados de desarrollo celular. Los patrones digitales fueron obtenidos previamente mediante un proceso de digitalización de imágenes biomédicas. El proceso de digitalización es complejo e involucra el empleo de un microscopio electrónico [ARÁ01].

El problema es tal que aún para el experto humano es difícil discernir el estado celular de los patrones, ya que los parámetros que definen cada clase de patrón de estado de desarrollo celular presentan características estadísticas propias que lo hacen confuso; esto es, los patrones de estados celulares por su propia naturaleza no son separables de forma sencilla. A esto también debe incluirse que, de forma natural los porcentajes de cada clase de desarrollo celular presentan un sesgo implícito; esto es, que proporcionalmente en la población total de patrones, los tipos de desarrollo celular no están representados de forma equitativa. De alguna forma es propiedad de las poblaciones naturales de células, que los estados de desarrollo celular tiendan a tener una proporción mayor a manifestarse en una determinada etapa de desarrollo que en otra.

## 1.2 MARCO DE ESTUDIO.

Antes de diseñar y entrenar una red neuronal, el experto en redes neuronales debe realizar un preprocesamiento a los datos disponibles que constituyen el conjunto experimental de datos del problema. Generalmente el experto mismo diseña estrategias para optimizar la separación del conjunto de datos experimentales, en un conjunto de entrenamiento y un



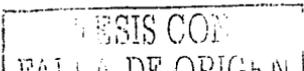
conjunto de validación[PRI96], [PRI00a], [SMI93]. Tal optimización es principalmente con objeto de elegir un conjunto de entrenamiento que sea suficientemente representativo, es decir, que no se presente inexactitud o exceso de aproximación al conjunto universo que modela la naturaleza propia del problema a resolver. También se debe cuidar que el tamaño del conjunto de entrenamiento no se traduzca en exceso de tiempo de entrenamiento de la red neuronal. Así la selección de un conjunto de entrenamiento tiene que obedecer dos criterios principales: representatividad y tamaño, y éstos son inversamente proporcionales; una muestra de entrenamiento óptima debe ser representativa y a su vez pequeña en cantidad. Usualmente el resto del conjunto de datos experimentales se toma como conjunto de validación; como el término lo indica, este conjunto sirve para validar el poder de generalización que ha alcanzado la red neuronal después del proceso de autoajuste con la muestra de entrenamiento. Es un hecho que en la gran mayoría de los problemas reales, es imposible contar completamente con todo el conjunto universo de datos del problema, inclusive es difícil contar con un conjunto de datos experimentales que sea representativo del conjunto universo que modela el problema[SMI93]. La prueba más importante de toda red neuronal se presenta cuando se van a procesar conjuntos nuevos de datos que no estaban incluidos dentro de la muestra experimental; es aquí donde se justifica la búsqueda de generalidad en redes neuronales, si bien la red neuronal puede dar buenos resultados con el conjunto de validación, esto no asegura que con un conjunto de datos nuevos esto suceda. Así, dado que la completa disposición de los datos experimentales regularmente está fuera del alcance del experto en redes neuronales, entonces la solución se propone desde otro enfoque, el problema de generalidad también se puede atacar modificando la arquitectura de la red neuronal.

La principal aplicación de una red neuronal es aproximar, de forma cercana, la superficie que describe una función objetivo, naturalmente deseamos que la aproximación no sea más compleja que la función objetivo misma. Las redes neuronales son aproximadores universales muy potentes, cada nodo en la red es un parámetro que incrementa la capacidad de aproximación de la red. El número de pesos determina los grados de libertad, más relevantes, con los cuales la red es capaz de ajustar los datos; limitar o incrementar el poder de la red neuronal significa calibrar el número y magnitud de los pesos. A su vez el número de pesos en la red es una función del número de neuronas. Generalmente el problema

mismo propone el número de neuronas de entrada y salida, dejando entonces la opción sobre las neuronas de las capas intermedias. El desarrollador de redes neuronales sabe que un número precario de neuronas intermedias resulta en bajo ajuste y un número excesivo de neuronas resulta en sobre ajuste. Además otro factor relacionado con el número de neuronas, es el tiempo, cada neurona consume una fracción del tiempo de entrenamiento y en consecuencia una fracción de tiempo de cómputo. En concreto el número de neuronas y capas intermedias en una red neuronal puede ser la diferencia entre una red exitosa y una red mala.

Un número de técnicas empíricas se han desarrollado y justificado para establecer y calibrar apropiadamente la arquitectura de una red neuronal, dado un problema particular; al respecto no existen fórmulas conocidas, existen algunas reglas, la mayoría basadas en la práctica y se toman decisiones sobre la marcha o posteriormente al entrenamiento de la red neuronal, lo cual degrada su eficiencia[CHI96]. Existen algunas técnicas de podado y crecimiento neuronal basadas en: dada una arquitectura base establecida, ir eliminando o incorporando, neuronas y/o capas intermedias, dependiendo de los requerimientos del problema, con el objetivo de mejorar el desempeño de la red neuronal. Sin embargo estas técnicas tienden a requerir mucho conocimiento acerca de la actividad de cada neurona individual y es necesario realizar estimaciones del problema en particular, lo cual no siempre esta muy disponible, y sobre todo, estas técnicas de poda y crecimiento neuronal están limitadas, dada su naturaleza, a explorar un número muy limitado de familias de configuraciones neuronales.

Hasta aquí sólo se ha mencionado la problemática en cuanto a la arquitectura neuronal, vista como un problema de construcción, no se ha hecho mención a la parte funcional de la arquitectura neuronal, que bien es una parte importante del diseño de redes neuronales ya que se deben hacer consideraciones acerca de la selección de las funciones de activación y ser muy cuidadoso en su planteamiento. Se sabe que las funciones de activación deben cumplir las condiciones de ser continuas y derivables, para que los algoritmos basados en retropropagación (Backpropagation) puedan seguir la trayectoria que establece el gradiente descendente en la etapa de retropropagación; esto limita a una familia particular de funciones. El papel de las funciones clásicas de activación, que si bien en la mayoría de los casos han cumplido bien su objetivo, deja a un lado, un número importante de funciones



que dadas las características especiales de un problema determinado bien podrían ser útiles como características especiales de diseño y funcionamiento.

Ahora bien, conociendo a grandes rasgos la problemática a la que se enfrenta el experto en redes neuronales para encontrar una arquitectura de red óptima, podemos decir que en muchos casos la etapa de diseño se traslapa con la etapa de entrenamiento. También se deben tomar consideraciones importantes que están relacionadas directamente con la arquitectura de la red neuronal. Para empezar, la red neuronal es sensible a las condiciones iniciales de entrenamiento. Se han desarrollado varias técnicas, algunas basadas en la estadística clásica, pero éstas suponen conocimiento a priori sobre el comportamiento de la red neuronal y sobre sus matrices de pesos iniciales, pero en la mayoría de los casos esto no es comprobable directamente, sino hasta que la red ya empieza a ser entrenada con algún algoritmo neuronal, inclusive, son necesarios varios entrenamientos para establecer una comparación entre condiciones iniciales diferentes[CH196].

Actualmente el diseño, entrenamiento y calibración de redes neuronales se realiza mediante: variantes del algoritmo de Backpropagation, Algoritmos Genéticos, Programación Genética y el modelo de Kohonen con sus variaciones[LEO98]. Todas las técnicas tienen sus ventajas y desventajas.

Las Técnicas Evolutivas formulan el entrenamiento y diseño de redes neuronales como un problema de búsqueda en el espacio de pesos y arquitecturas, respectivamente, donde cada punto representa un conjunto de pesos o una arquitectura, susceptible probabilísticamente de ser la solución óptima para el problema en cuestión.

Los algoritmos genéticos son aplicados para mejorar algunas propiedades de las redes neuronales, su utilidad radica principalmente en, mediante una búsqueda probabilística, establecer un entrenamiento genético, es decir, mediante procesos de selección probabilística, intercambio y modificación aleatorios de los valores de los pesos, encontrar matrices de pesos óptimas. Pero estos llegan sólo hasta esa parte, dada la limitada estructura cromosómica que es capaz de manejar el algoritmo genético, que en general proponen una codificación binaria para establecer el genotipo.

La Programación Genética, como directamente se define, es una extensión de los algoritmos genéticos que permite manipular estructuras cromosómicas más complejas. Esta propiedad permite traducir por completo, en una estructura cromosómica, la arquitectura

de una red neuronal, incluyendo funciones de activación y otros parámetros propios de la arquitectura neuronal. Sin embargo esta capacidad paga un precio cuando, dada la tendencia original de las técnicas evolutivas, se codifican binariamente las estructuras cromosómicas, así el manejo computacional se vuelve ineficiente.

A continuación se presentan algunos trabajos interesantes, todos ellos son propuestas integradoras de las redes neuronales y la computación evolutiva.

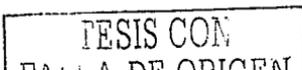
### 1.3 ANTECEDENTES

- **A Real-Coded Genetic Algorithm for Training Recurrent Neural Networks**  
A., Blanco, M. Delgado, M.C. Pegalajar

En este trabajo se presenta y justifica una codificación real vectorial para representar redes neuronales recurrentes en una estructura cromosómica neuronal real sencilla. Sólo toma en cuenta el uso de un vector con valores reales para representar la matriz de pesos de una red neuronal de arquitectura fija. La función objetivo que propone es la expresión del error entre los valores de salida de la red y los valores target. Así el entrenamiento genético que propone se reduce a un problema de combinatoria, es decir, el algoritmo genético busca la matriz de pesos que mejor reduzca el error entre las salidas de la población de cromosomas neuronales reales y los valores target del problema a resolver. Presenta una variedad interesante de implementaciones de operadores genéticos, distintos tipos de mutaciones y cruza propios para la codificación real y compara los resultados obtenidos con el entrenamiento neurogenético utilizando varias combinaciones de estos operadores genéticos.

- **Discovery of Backpropagation Learning Rules Using Genetic Programming**  
Amr Radi and Ricardo Poli

Este trabajo es interesante porque hace uso de la Programación Genética para desarrollar evolutivamente estructuras cromosómicas que representan expresiones matemáticas, y mediante operadores genéticos encontrar ecuaciones de ajuste de pesos neuronales que optimicen las reglas de ajuste que propone el algoritmo de backpropagation. Los autores proponen dos pequeños conjuntos, uno de operadores aritméticos, básicamente suma, resta y multiplicación, también conocidos como no terminales; el otro conjunto lo forman



operandos o terminales, el cual incluye las salidas de las neuronas, los pesos asociados, los términos de error, los términos de tendencia, las tasas de aprendizaje y algunas constantes numéricas. Con ambos conjuntos, terminales y no terminales, en un principio aleatoriamente y posteriormente evolutivamente, se desarrollan estructuras cromosómicas que se decodifican en árboles sintácticos para establecer una evaluación mediante una función objetivo que utiliza como criterios a optimizar el número de epochs de entrenamiento y el error cuadrático promedio. Sin embargo, la arquitectura neuronal es fija, no toma en cuenta las funciones de activación para establecer mejoras en el aprendizaje, y no hace consideración alguna de las condiciones iniciales de entrenamiento.

- **Evolving Neural Networks Structures by means of Genetic Programming**  
Wolfgang Golubsky and Tomas Feuring.

Este artículo presenta una técnica que simultáneamente construye topologías y entrena redes neuronales mediante programación genética. Propone una codificación cromosómica de red neuronal basada en la sintaxis de LISP, mediante listas, la decodificación es una representación de árbol de sintaxis o grafo dirigido. El trabajo propone varios parámetros genéticos basados en los resultados obtenidos. El operador genético de cruzamiento realiza mecanismos de intercambio de características topológicas únicamente en las neuronas de las capas intermedias y en los pesos asociados a éstas; ofrece dos propuestas del operador de cruzamiento para mejorar la topología de la red neuronal, pero no toma en cuenta el operador de mutación. La función de aptitud para seleccionar arquitecturas óptimas toma en cuenta un criterio restrictivo en el tamaño de la red neuronal.

- **A New Evolutionary System for Evolving Artificial Neural Networks**  
Xin Yao and Yong Liu

Los autores contribuyen a la técnica neuroevolutiva con un completo y novedoso enfoque para entrenar y desarrollar redes neuronales. A diferencia del trabajo de Golubsky y Feuring, este enfoque se basa en el operador genético de mutación como agente evolutivo único para optimizar la topología y simultáneamente entrenar a las poblaciones de redes neuronales. Lo novedoso aquí es que el operador de mutación es prácticamente por sí mismo un entrenamiento backpropagation con tasa de aprendizaje adaptable, número de

epochs predeterminado conjugado con un proceso de poda y crecimiento neural aleatorio. Básicamente si al final de esta mutación-entrenamiento neuronal la red entrenada (red hija) resulta en una reducción del error significativa con respecto al error de la red antes del entrenamiento neuronal (red padre), la red hija obtiene un valor de aptitud mayor y por consiguiente es más probable preservar su herencia al ser seleccionada como red padre en la siguiente generación. Pero aquí no termina el operador de mutación propuesto, una vez que las redes hijas (redes entrenadas con backpropagation) sustituyen a las redes padres, a cada una se aplica un proceso de poda y crecimiento en las capas intermedias, mediante una distribución uniformemente aleatoria, y se reentrena con backpropagation con fin de obtener descendencia de redes modificadas topológicamente y robustamente entrenadas. Una vez finalizada la operación de mutación, el algoritmo evolutivo se itera un determinado número de generaciones. La asignación de la aptitud global es una función determinada por una expresión basada en el error cuadrático promedio. El método propuesto en este artículo, se reporta probado en una amplia variedad de aplicaciones.

Sin embargo, nuevamente la estructura cromosómica que se presenta en este trabajo está basada en una forma vectorial codificada, más exacto una matriz de adyacencias, es decir, se tiene que pasar de genotipo a fenotipo; además, las redes neuronales que genera son de estructura muy sencilla, no muy útiles cuando el problema a resolver es multiobjetivo, donde se requieren arquitecturas con más de dos neuronas de salida. También es deseable incluir en el cromosoma términos de tendencia y variedad de funciones de activación y no usar solo un tipo de función de activación, la sigmoide para todas las neuronas de la población de redes neuronales como se propone en este trabajo. Además el autor plantea la mutación como el único operador evolutivo, no explota por completo las características hereditarias de la población de redes neuronales en las sucesivas generaciones del programa evolutivo.

Reflexionando sobre los trabajos anteriores, en este trabajo vamos a integrar aquellas ideas que son relevantes e incluiremos algunas otras que consideramos optimizan al diseño y entrenamiento de CNAs. Esto nos conduce a definir los objetivos particulares que serán las directrices principales de esta tesis.

TESIS CON  
FALLA DE ORIGEN

#### 1.4 OBJETIVOS

- Integrar las técnicas evolutivas y neuronales para ofrecer una metodología en el diseño de redes neuronales ofreciendo una codificación cromosómica neuronal real directa, fenotípica, como se muestra en la figura 1.2.
- Proponer e implementar un conjunto de operadores genéticos matriciales propios para manipular y modificar la estructura cromosómica real directa, fenotípica, que representa la arquitectura neuronal.
- Analizar la influencia del cambio de la arquitectura de la red y sus consecuencias en el entrenamiento y desarrollo neuroevolutivo.
- Implementar una propuesta propia basada en técnicas estadísticas, algoritmos evolutivos y toma de decisiones multiobjetivo para el desarrollo y evaluación de redes neuronales clasificadoras.
- Examinar las consecuencias de manipular las combinaciones de funciones de activación, en la configuración y entrenamiento genético de una red neuronal clasificadora.

La estructura ilustrada en la figura 1.2 es nuestra estructura cromosómica genérica, es decir es la estructura general que tendrán nuestros CNAs, cabe ser específico en este momento, sobre la estructura de la figura 1.2., el número de capas es un factor constante y se determina en tres capas: de entrada, una capa intermedia y capa de salida. El número de neuronas de entrada y salida están determinadas por el problema a resolver. El número de neuronas en la capa intermedia es un parámetro a optimizar, es decir, nuestras redes neuronales tendrán la capacidad de plasticidad en la capa intermedia; esta plasticidad, por supuesto, estará acotada por un valor mínimo y un valor máximo de neuronas en la capa intermedia. Como funciones de activación,  $f$ , se tomarán las clásicas funciones sigmoideas y un conjunto reducido de funciones de la familia de las funciones derivables. El objetivo de utilizar una estructura cromosómica fenotípica obedece a que en la mayoría de los

trabajos reportados en el área, proponen una codificación cromosómica binaria, lo cual obliga a realizar una decodificación de genotipo a fenotipo. Lo que resulta en un problema de permutación o, como la literatura lo reporta, un problema de convención de contenido; lo que significa que redes con distinto genotipo (codificación) tienen comportamiento equivalente en fenotipo (decodificación). Este problema no sólo hace ineficiente el proceso de evolución de los CNAs, también hace más difícil producir descendientes altamente adaptados mediante la operación de cruzamiento. Además que el problema de convención también contribuye al fenómeno de la deriva genética, el cual analizaremos con detalle.

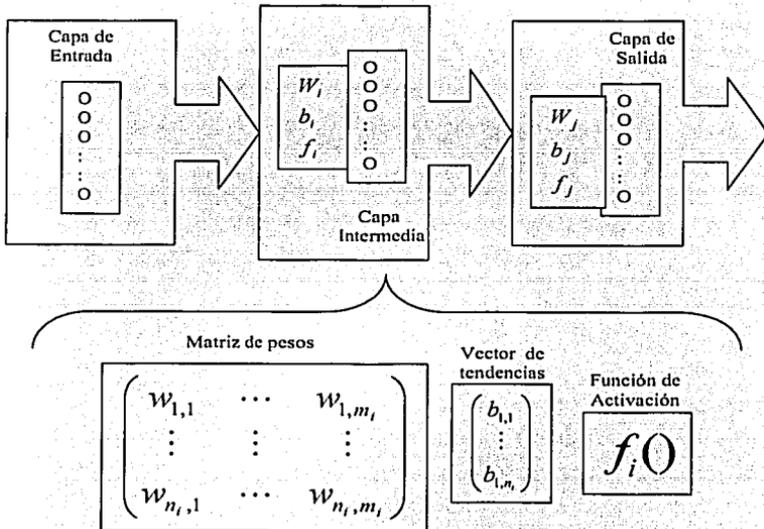


Figura 1.2 Estructura cromosómica que representa a una red neuronal y detalle esquemático de la subestructura que compone una capa

## 1.5 APORTACIONES

- El trabajo contribuye con una metodología integral propia para el problema general de desarrollo y entrenamiento de clasificadores neuronales artificiales basada en la programación evolutiva multiobjetivo.
- Se desarrollan y prueban operadores genéticos propios para el desarrollo y entrenamiento de clasificadores neuronales artificiales.
- Se analiza y prueba una medida alternativa multiobjetivo para la evaluación del desempeño de clasificadores neuronales artificiales.
- Se propone y prueba una solución alternativa para el fenómeno de deriva genética[DEB89].
- El trabajo reporta resultados experimentales para el problema del diseño y entrenamiento de clasificadores neuronales artificiales de patrones de etapas de desarrollo celular.
- Toda la herramienta de programación desarrollada en este trabajo para la prueba de las ideas presentadas mediante la verificación de los resultados obtenidos representa un sistema completo para el diseño entrenamiento y optimización de sistemas neuronales artificiales basado en técnicas de la programación evolutiva.

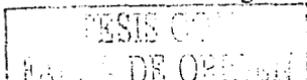
## 1.6 CONTENIDO DE LA TESIS

En el capítulo uno presentamos las características particulares del problema de clasificación de patrones de etapas de desarrollo celular y las condiciones en las cuales se encuentra la muestra experimental de datos. Hacemos una revisión de algunas de las técnicas básicas con las que actualmente se realiza el entrenamiento y desarrollo de clasificadores neuronales, analizamos sus ventajas y desventajas. Tomamos en cuenta la problemática de la deriva genética y hacemos una crítica a la expresión del error cuadrático medio y con base a ésta justificamos y presentamos las definiciones elementales y elementos básicos de

la toma de decisiones multicriterio. Finalmente establecemos y justificamos la configuración básica de Clasificador Neuronal Artificial (CNA) que vamos a utilizar en todo el desarrollo del trabajo.

En el capítulo dos presentamos y justificamos la serie de técnicas específicas que integran nuestra propuesta metodológica para el diseño y entrenamiento de clasificadores neuronales artificiales. Explicamos a detalle las partes operacionales de nuestra propuesta de solución y visualizamos de forma funcional su factibilidad de conjunción. Presentamos y ejemplificamos una técnica clásica para el muestreo representativo de patrones de etapas de desarrollo celular, al mismo tiempo analizamos las singularidades de la muestra experimental de patrones de desarrollo celular. Posteriormente presentamos de forma explícita la estrategia general para el diseño y entrenamiento de CNAs y definimos las variables de diseño y entrenamiento y simultáneamente delimitamos sus rangos de operación. Esto nos conducirá a la definición final de la estructura cromosómica con la cual trabajará nuestra metodología de diseño y entrenamiento. Esta definición nos obligará a definir un número finito de variables de diseño, que operen en un rango delimitado de valores. Con la determinación de la estructura cromosómica de trabajo, en seguida determinamos nuestra variante del algoritmo genético simple, la cual define familias independientes de CNAs. La justificación de nuestra propuesta particular, la determinamos por medio de la definición específica de los operadores genéticos de migración, cruzamiento y mutación que hemos diseñado propiamente para nuestra metodología. Dentro de los operadores genéticos, tomamos y justificamos el uso de una técnica alterna a la técnica de pareto para el ordenamiento de los elementos dentro de las familias de la población de CNAs. Finalmente presentamos la conveniencia de utilizar el concepto de matriz de confusión como medida alterna para la evaluación del desempeño de los CNAs, y utilizamos tal concepto para la definición de la función multiobjetivo, la cual contiene una combinación de heurística estadística y valoraciones cuantitativas que nos dan métricas cuantitativas multiobjetivo para la evaluación de los CNAs obtenidos mediante nuestra metodología de diseño y entrenamiento.

El capítulo tres muestra los resultados obtenidos, para el problema específico de clasificación de patrones de estados de desarrollo celular. La combinación de técnicas que forman nuestra metodología son probadas sobre un conjunto experimental de patrones



específico que tiene varias singularidades y lo hacen muy adecuado para ser objeto de prueba de la metodología de diseño y entrenamiento automático de CNAs propuesta. En el capítulo cuatro se presentan las conclusiones finales de este trabajo y se muestran algunas alternativas futuras de investigación.

TESIS CON  
FALLA DE ORIGEN

## **CAPÍTULO UNO**

### **PROBLEMÁTICA DEL DISEÑO Y ENTRENAMIENTO DE UN CLASIFICADOR NEURONAL ARTIFICIAL (CNA) PARA EL RECONOCIMIENTO DE PATRONES DE ETAPAS DE DESARROLLO CELULAR.**

#### **1.1 DISTRIBUCIONES ESTADÍSTICAS DE LOS PATRONES DE DESARROLLO CELULAR**

Este trabajo es continuación de un proyecto para el diseño y entrenamiento de un Clasificador Neuronal[PR197]. El objeto de tal proyecto consistió en diseñar un Clasificador Neuronal Artificial (CNA), el cual semiautomatiza un sistema que reconoce y cuenta células en diferentes estados de desarrollo. Tal proyecto anteriormente fue desarrollado con un clasificador diseñado por un experto en redes neuronales y entrenado con técnicas clásicas neuronales, específicamente utilizando variantes del algoritmo de Backpropagation. En este trabajo el enfoque es distinto, nosotros vamos hacer uso de las técnicas evolutivas para encontrar las matrices de pesos, los vectores de términos de tendencia, las funciones de activación y la configuración estructural óptima del CNA. Así solucionamos simultáneamente el problema de diseño del CNAs y el problema de clasificación de patrones de estados de desarrollo celular.

En la química y la medicina, las tasas de proliferación celular, es decir, los porcentajes de células en diferentes etapas de diferenciación, son indicadores estrechamente relacionados con el diagnóstico y pronóstico del cáncer, entre otras enfermedades. Cada tipo de células se divide con una periodicidad óptima para el correcto mantenimiento y funcionamiento del organismo; hay células que se dividen con gran frecuencia, como las que recubren la pared del intestino, y otras que lo hacen muy esporádicamente o incluso nunca, como las neuronas y la mayoría de las células nerviosas. La proliferación celular es la base del desarrollo de los organismos, según un proceso de división celular, en el que cada tipo celular se divide cuando y cuanto debe. En el momento en que este proceso natural se altera, ocurre el primero de los procesos que origina un cáncer: el crecimiento descontrolado de un grupo de células. El ciclo celular es un conjunto ordenado de eventos que culmina con el crecimiento de la célula y la división en dos células hijas[**DAR00**]. Las células que no se encuentran en proceso de división no se consideran que estén en el ciclo

TESIS CON  
FALLA DE ORIGEN

# PAGINACIÓN DISCONTINUA

celular. La mitosis es el proceso de división celular por el cual se conserva la información genética contenida en sus cromosomas, que pasa de esta manera a las sucesivas células a que la mitosis va a dar origen. La mitosis es, por sí misma, un verdadero proceso de multiplicación celular que participa en el desarrollo, el crecimiento y la regeneración del organismo. El proceso tiene lugar por medio de una serie de operaciones sucesivas que se desarrollan de una manera continua, y que para facilitar su estudio han sido separadas en varias etapas. Véase figura 1.1.

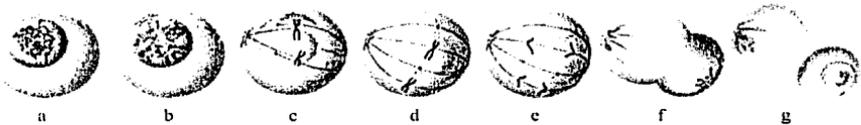


Figura 1.1. Proceso de la Mitosis: (a) Interfase, (b) Profase, (c) Prometáfase, (d) Metafase, (e) Anafase, (f) Telofase, (g) Citocinesis.

- **Interfase.-** La célula está ocupada en la actividad metabólica preparándose para la mitosis (las próximas cuatro fases que conducen e incluyen la división nuclear). Los cromosomas no se disciernen claramente en el núcleo, aunque una mancha oscura llamada nucleolo, puede ser visible. La célula puede contener un par de centriolos (o centros de organización de microtúbulos en los vegetales) los cuales son sitios de organización para los microtúbulos.
- **Profase.-** La cromatina en el núcleo comienza a condensarse y se vuelve visible en el microscopio óptico como cromosomas. El nucleolo desaparece. Los centriolos comienzan a moverse a polos opuestos de la célula y fibras se extienden desde los centrómeros. Algunas fibras cruzan la célula para formar el huso mitótico.
- **Prometáfase.-** La membrana nuclear se disuelve, marcando el comienzo de la prometáfase. Las proteínas se adhieren a los centrómeros creando los cinetócoros.

Los microtúbulos se adhieren a los cinetócoros y los cromosomas comienzan a moverse.

- **Metafase.-** Fibras del huso alinean los cromosomas a lo largo del medio del núcleo celular. Esta línea es referida como, el plato de la metafase. Esta organización ayuda a asegurar que en la próxima fase, cuando los cromosomas se separan, cada nuevo núcleo recibirá una copia de cada cromosoma.
- **Anafase.-** Los pares de cromosomas se separan en los cinetócoros y se mueven a lados opuestos de la célula. El movimiento es el resultado de una combinación de: el movimiento del cinetócoro a lo largo de los microtúbulos del huso y la interacción física de los microtúbulos polares.
- **Telofase.-** Los cromátidos llegan a los polos opuestos de la célula, y nuevas membranas se forman alrededor de los núcleos hijos. Los cromosomas se dispersan y ya no son visibles bajo el microscopio óptico. Las fibras del huso se dispersan, y la citocinesis o la partición de la célula puede comenzar también durante esta etapa.
- **Citosinesis.-** En células animales, la citocinesis ocurre cuando un anillo fibroso compuesto de una proteína llamada actina, alrededor del centro de la célula se contrae pellizcando la célula en dos células hijas, cada una con su núcleo. En células vegetales, la pared rígida requiere que una placa celular sea sintetizada entre las dos células hijas.

Las preparaciones de muestras de células para análisis médico o químico, presentan cientos de células, actualmente existe un sistema de adquisición de imágenes mediante un microscopio óptico, al final de esta adquisición el experto identifica y cuenta cientos de células y reporta el censo de formas mitóticas de entre los núcleos de la interfase, es decir, realiza una clasificación entre metafase y núcleos[ARÁ01], [COR98], [GAR95]. Este es el propósito del experto en biología celular y del clasificador neuronal, determinar los porcentajes de estados de desarrollo de las células, mediante un conjunto de patrones obtenido con un proceso digital de imágenes biomédicas. Cada patrón está formado por cinco parámetros[GAR95]:

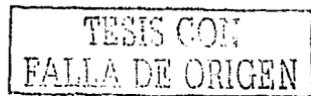
TESIS CON  
FALLA DE ORIGEN

- Forma es una característica de gran utilidad ya que según la etapa de desarrollo en que se encuentre la célula, ella posee una forma más circular que en otras etapas. Por ejemplo los núcleos tienen una forma casi circular y el sistema de adquisición de características a estas células les asigna un valor muy cercano a 1 en su factor forma.
- Área, esta característica, junto con la de Forma, es una de las primordiales que toma en consideración el experto para distinguir una célula en franca etapa de mitosis, de una célula que apenas ha iniciado la etapa de interface. Las células en etapa de interface poseen un rango típico de tamaño. El valor de tamaño que distingue a una célula en plena etapa de mitosis, de una célula en etapa próxima a iniciar la mitosis, se traduce a un valor de área en píxeles.
- Nivel de gris es una característica importante para el experto, puede significar una medida de la textura de un objeto. Específicamente la manipulación digital del nivel de gris en los contornos de las formas celulares permite un mejor aislamiento de los contornos que pueden representar distintas etapas de desarrollo celular.
- Desviación estándar del nivel de gris La forma en que el ojo humano reconoce objetos, como las mitosis, es por medio de las texturas. La textura da una sensación de relieve. Los niveles de gris, en una imagen digital, se traducen de alguna forma en profundidades y elevaciones. De la imagen de una mitosis o de un núcleo, contenido en un contorno definido, se obtiene la desviación estándar del nivel de gris, como indicador de relieve que puede caracterizar una etapa de desarrollo celular.
- Dep. Hoyos Es el valor medio del radio de la profundidad-amplitud de una depresión en la imagen de un objeto. Este valor es obtenido mediante un barrido en línea horizontal, realizado mediante un procesamiento digital de adquisición imágenes[COR98].

En el proyecto previo, se realizó una estadística descriptiva con objetivo de conocer las distribuciones de cada parámetro y obtener los principales valores estadísticos para normalizar los parámetros de los patrones, con el fin de adecuarlos para diseñar un clasificador neuronal artificial, el cual se diseñó utilizando funciones de activación del tipo sigmoide[PR196].

El número de parámetros que caracterizan a los patrones de etapas de desarrollo celular, fue seleccionado por los expertos que a su consideración son suficientes para diferenciar las tres clases de patrones celulares: Basura el cual es la mezcla de material celular sobrante con diversos reactivos químicos; Mitosis y Núcleos que son dos estados de desarrollo de la célula. Distinguir de entre los tres tipos de patrones no es sencillo, aún para el experto humano, ya que como se observó en los trabajos previos, existe redundancia, ambigüedad entre parámetros como lo demuestran las distribuciones de cada parámetro, ya que en ningún caso se tiene una separación bien definida del estado de la célula[PR100a], [PR100b], [GAR95]. La muestra experimental tiene 900 patrones, es la muestra experimental más reciente de patrones de células que fue proporcionada por los expertos biomédicos, la cual consta de 182 patrones asociados a mitosis, 331 patrones clasificados como núcleos y 387 patrones correspondientes a basuras.

Observemos la figura 1.2, en ella se muestran las distribuciones relativas de los elementos de la muestra para cada clase tomando en cuenta únicamente el parámetro Forma; los valores de la media son: 0.7514, 0.9174, 0.7396 y los de la mediana son: 0.8126, 0.9288, 0.7801 para mitosis núcleos y basura, respectivamente. Esto nos indica que para el parámetro Forma los valores tienden a estar por arriba del valor 0.7 en las tres clases. Pero que como lo indica el valor de la media para las clases mitosis y basuras, el punto medio de la distribución en ambos casos esta alrededor de 0.7, lo que no sucede para núcleos cuyo punto medio esta cerca de 0.9. Pero como se nota en la figura 1.2 y lo corroboran los valores de las medias, los puntos de equilibrio son distintos para los tres casos[WON98]. Los tres hechos nos indican que para el parámetro Forma existe un compromiso, no muy definido, entre mitosis y basuras lo cual hace que el parámetro Forma no sea un claro factor discriminante.



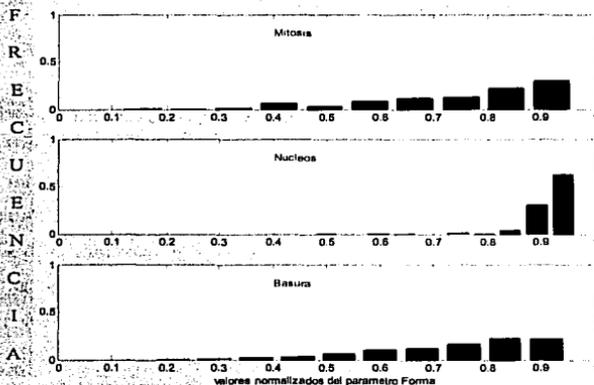


Figura 1.2 Gráficas de las distribuciones relativas de los valores del parámetro Forma para cada clase de patrones de la muestra experimental

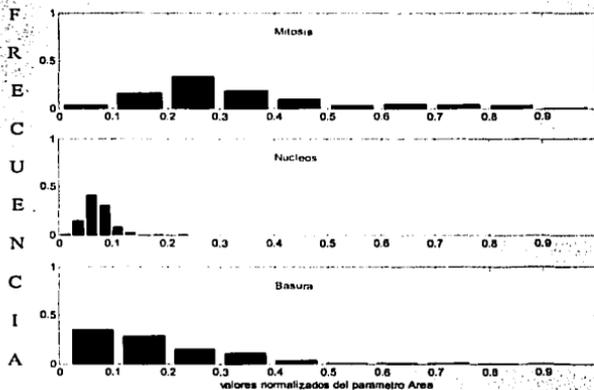


Figura 1.3 Gráficas de las distribuciones relativas de los valores del parámetro Área para cada clase de patrones de la muestra experimental

La figura 1.3 corresponde a las gráficas de las distribuciones relativas de los valores para el parámetro Área, en ella podemos notar que para mitosis, núcleos y basuras los valores de las medianas y medias son 0.3392, 0.0718, 0.1908 y 0.2767, 0.0693, 0.1497 respectivamente. Aquí a diferencia que para la Forma el compromiso parece estar entre núcleos y basura, ya que tanto su punto medio como su punto de equilibrio son cercanos, pero tal compromiso no es tan marcado como en el caso de Forma.

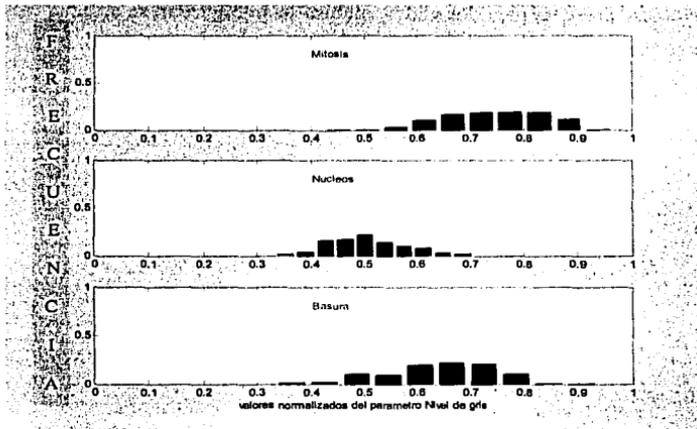


Figura 1.4 Gráficas de las distribuciones relativas de los valores del parámetro Nivel de Gris para cada clase de patrones de la muestra experimental

En la figura 1.4 observamos las distribuciones relativas para el parámetro Nivel de Gris. Los valores correspondientes para mitosis, núcleos y basura son: para las medias 0.7446, 0.5065, 0.6400 y para las medianas 0.7450, 0.5000, 0.6500. Tales valores, aunque no claramente a diferencia de los parámetros Forma y Área, son más diferenciables entre los tres. También la figura 1.4 sugiere de alguna manera más distinción en los puntos medios y los puntos de equilibrio de las tres gráficas contenidas.

Los valores 0.2502, 0.4624 y 0.3352 corresponden a los valores de las medias del parámetro Dep. Hoyos para mitosis, núcleos y basura respectivamente. Así como también respectivamente, a cada clase del mismo parámetro, corresponden los valores de las

medianas 0.2318, 0.4607 y 0.3369. Las gráficas de las distribuciones relativas para este parámetro se muestran en la figura 1.5. Nótese que Dep. Hoyos observa la misma diferenciabilidad que el parámetro Nivel de gris. La diferenciabilidad a la que nos referimos, en ambos parámetros, es en cuanto a que si bien los puntos de equilibrio y los puntos medios de las distribuciones para cada clase no están bien definidos, en ambos casos solo existe una décima de diferencia, pero por lo menos tal diferenciación ocurre en las tres clases; no como se presenta en los parámetros Área y Forma, donde la similitud en los valores estadísticos se presenta para dos clases, incluso la ambigüedad se acentúa más en el parámetro Forma que en el parámetro Área como analizamos anteriormente.

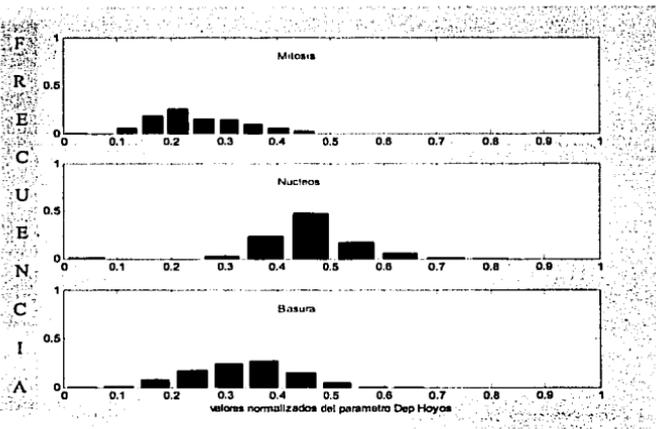


Figura 1.5 Gráficas de las distribuciones relativas de los valores del parámetro Dep. Hoyos para cada clase de patrones de la muestra experimental

Del análisis estadístico para los cuatro parámetros anteriores surge la pregunta ¿por qué no incluir la moda en el análisis? Si bien la moda nos da idea de cual es el valor más frecuente dentro de una muestra experimental, no es tan buena medida de tendencia como las medias y las medianas, ambos son estimadores más consistentes de la tendencia de la muestra. Como vemos en la figuras 1.4 y 1.2 tenemos distribuciones trimodales y bimodales en los

parámetros Nivel de Gris y Forma respectivamente, lo cual indica que la frecuencia para esos dos parámetros ocurre más de una vez y por tanto como medida de tendencia central para esta muestra particular de patrones de etapas de desarrollo celular, tomar las modas es poco útil.

Finalmente analicemos el parámetro Desv. Est. del Nivel de gris, cuyas distribuciones están en la figura 1.6.

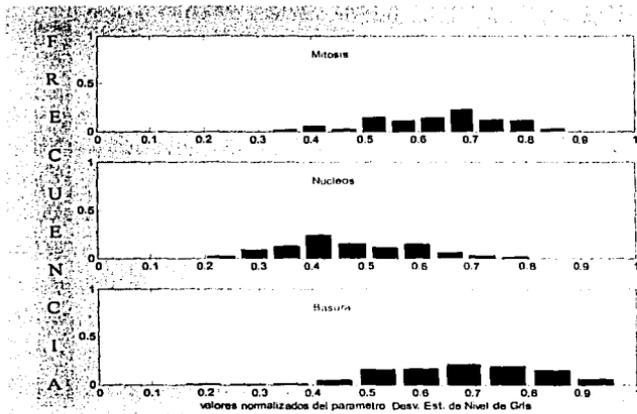


Figura 1.6 Gráficas de las distribuciones relativas de los valores del parámetro Desv. Est. del Nivel de gris para cada clase de patrones de la muestra experimental

El parámetro Desv. Est. del Nivel de gris, es el último parámetro de los cinco que conforman a los patrones de etapas de desarrollo celular. Veamos cuales son los valores de las medias y las medianas para cada clase. Los valores de las medianas para mitosis núcleos y basuras son: 0.6444, 0.4556 y 0.6889 y también respectivamente los valores de las medias son: 0.6363, 0.4684 y 0.6791. Comparando los valores de las medianas podemos decir que, nuevamente los núcleos son más diferenciables de entre las mitosis y las basuras. La confirmación de tal conclusión proviene también de la comparación de los valores de las

medias que nos indican que el punto de equilibrio de las distribuciones de mitosis y basura son muy cercanos, alrededor de 0.6.

Nótese que este análisis estadístico que hemos realizado ha sido mediante distribuciones estadísticas relativas y con valores normalizados de parámetros, ya que de esta forma es más sencillo realizar comparaciones entre los valores estadísticos de mitosis núcleos y basuras; debido a que se elimina el problema de considerar distintas escalas, el cual en el caso de la muestra de patrones de desarrollo celular si es un factor que se debe tomar en cuenta, porque entre los cinco parámetros no existe homogeneidad en las magnitudes, además que las proporciones de patrones de mitosis, núcleos o basura no son homogéneas como se verá en la siguiente sección.

## **1.2 SESGO IMPLÍCITO EN LA POBLACIÓN EXPERIMENTAL DE PATRONES**

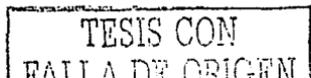
En el desarrollo de clasificadores previos, se ha contado con diversas poblaciones de patrones, como mencionamos anteriormente. En este caso tenemos una población experimental de 900 patrones, es la última población de patrones de desarrollo celular que fue proporcionada por el experto biomédico. La muestra posee 182 patrones asociados a mitosis, 331 patrones pertenecientes a núcleos y 387 patrones de la clase basuras. Debido a que en las muestras celulares reales, los expertos encuentran proporciones del orden de 8000 basuras, 40 mitosis y 2000 núcleos, entonces existe un sesgo implícito en todas las muestras celulares que manejan los expertos en los experimentos reales[ARÁ01], [COR98]. Éste, obviamente no es el caso de la muestra de patrones de estados celulares que tenemos disponible, es decir, no tiene un sesgo tan marcado según las proporciones que se presentan en las muestras celulares reales, pero aunque en menor porcentaje el sesgo existe, porque proporcionalmente tenemos que aproximadamente: 20% de los patrones pertenecen a la clase mitosis, 37% de los patrones son de la clase núcleos y el resto 43% de los patrones corresponden a la clase basuras. Estos porcentajes dispares corroboran de alguna forma que la muestra experimental que poseemos, es una muestra que fue tomada de manera tal que conserva, en cierta medida, las características propias de las poblaciones de células reales.

La prueba importante de todo CNA se presenta cuando va a clasificar muestras nuevas de datos, muy probablemente las muestras nuevas de patrones de desarrollo celular tendrán los

sesgos correspondientes a las clases con mayoría representativa de elementos, las basuras y los núcleos. Este sesgo natural tiene una consecuencia en el entrenamiento del CNA; los métodos de entrenamiento neuronales en general y el backpropagation en particular, son sensibles a las características estadísticas de las muestras de entrenamiento. Es preferible que la representatividad de todas las clases dentro de la muestra experimental tenga proporciones homogéneas, para que cada clase de la muestra tenga la misma influencia sobre el CNA; esto es, porque es más probable que elementos pertenecientes a clases proporcionalmente menores sean confundidos o incluidos dentro de los elementos que pertenecen a clases proporcionalmente mayores[SMI93]. La forma directa de evitar esta condición es incluir en la muestra que va a entrenar el CNA un número igual de patrones de cada clase, lo cual asegura la igualdad de representatividad proporcional. Pero si esta situación de sesgo proporcional es propia de las muestras reales de desarrollo celular, no debemos provocar ajustar el problema a la solución, sino contrariamente, que la solución sea adaptable a las características propias del problema. Por tal motivo, la técnica que implementemos debe ser capaz de lidiar con este sesgo natural. Tal objetivo lo atacaremos mediante las técnicas multiobjetivo.

Las técnicas multiobjetivo nos auxilian en la toma de decisiones, respetando siempre las características estadísticas de la muestra experimental, pero evitando que la tendencia hacia las clases núcleos y basura, que son las mayormente representadas, provoquen que la clasificación se sesgue. Ya que si bien las mitosis tienen una menor representación proporcional dentro de las poblaciones celulares reales, para los expertos es más importante que se clasifiquen mejor las mitosis que los núcleos y las basuras, porque el índice mitótico, el porcentaje de células en mitosis del total de células, puede ser un factor muy importante para diagnosticar y predecir algunos padecimientos tan graves como algunos tipos de cáncer[SUO97]. Es útil también en la prescripción de tratamientos delicados como la inmunosupresión. Así que, podemos decir que existe una jerarquía de importancia de clasificación, mitosis, núcleos y basuras, en ese orden. Así, es claro que todas las características estadísticas propias de la muestra experimental, deben ser tomadas en cuenta para dirigir el entrenamiento y diseño del CNA.

El análisis estadístico anterior es útil para tomar en consideración todas las dimensiones individuales del problema de clasificación de patrones de desarrollo celular. Al analizar las



características de la distribución de cada parámetro, nos adentramos en la problemática del diseño de un CNA que pueda clasificar correctamente todos los patrones de desarrollo celular. Ya que entrenar un CNA es un ejercicio de modelado estadístico, cuya meta es encontrar una ecuación que capture las generalidades y singularidades estadísticas de la muestra de patrones con que se entrene el CNA. Así, desde un punto de vista un CNA es sencillamente una ecuación, una función, cuyas variables independientes (variables de entrada), serán los patrones de etapas de desarrollo celular y el valor de la variable dependiente (el valor de salida), será el valor de la clase correcta para un patrón determinado. Pero dadas las características particulares del conjunto de datos experimentales, ¿Cómo vamos a encontrar un CNA óptimo que nos resuelva el problema de clasificación?, ¿De qué forma vamos a plantear la búsqueda simultánea de la topología y valores de pesos óptimos para el reconocimiento de patrones de etapas de desarrollo celular? A continuación se da una revisión genérica de las técnicas actuales para diseñar Redes Neuronales. De estas técnicas tomaremos los elementos e ideas principales para el planteamiento y refinamiento de nuestra propia propuesta de entrenamiento y desarrollo de un CNA para el reconocimiento de patrones de estados de desarrollo celular tomados a partir de imágenes biomédicas.

### 1.3 RETROPROPAGACIÓN

Los sistemas neuronales artificiales, o simplemente las Redes Neuronales se pueden definir como, el conjunto de técnicas computacionales (no convencionales) que toman ciertas características de la fisiología del cerebro como base para proponer nuevos modelos de procesamiento.

El algoritmo de retropropagación (*backpropagation*), uno de los más usados en Redes Neuronales, provee una manera de usar ejemplos de una función objetivo (función target) para ajustar los coeficientes (pesos) que hacen que una cierta función de mapeo se aproxime a la función objetivo tan cerca como sea posible[FRE97]. La función de mapeo utilizada en retropropagación es compleja. La manera más fácil de comprender esto es visualizarlo como el cálculo llevado a cabo por una red neuronal. Específicamente es el cálculo llevado a cabo por una red feedforward de tres capas completamente conectada, como se muestra en la figura 1.7. La red consiste de tres capas: capa de entrada, capa

intermedia y capa de salida. Cada neurona en la capa de entrada da a la red el valor de una variable independiente. Las neuronas en la capa intermedia hacen la mayor actividad de la red. Cada una de las neuronas de salida calcula una variable independiente.

De la red neuronal de la figura 1.7, podemos decir que es una red neuronal artificial, con cinco neuronas en la capa de entrada, tres neuronas en la capa intermedia o interna y tres neuronas en la capa de salida. Esta descripción nos da la configuración estructural de la red neuronal artificial. Pero esto no es todo, asociada a cada neurona existe un peso de conexión, un término de tendencia o bias y una función de activación, lo cual en conjunto plantea que para cada capa de la red neuronal, existe asociada una matriz de pesos, un vector de términos de tendencia y una función de activación. Es con esta descripción que se define completamente la arquitectura del CNA de la figura 1.7. Existen muchos tipos de arquitecturas de Redes neuronales. Para nuestro caso específico, el tipo de red neuronal *feedforward* es un buen mecanismo de cálculo para desarrollar las funciones de CNA de la muestra experimental de patrones de estados de desarrollo celular.

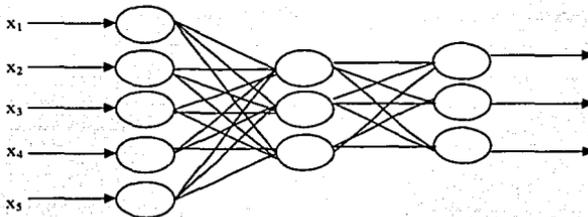


Figura 1.7 Red Neuronal Feedforward.

La red de la figura 1.7 está completamente conectada; esto indica que existen conexiones entre todas las neuronas en las capas adyacentes. Se tiene una conexión separada entre cada neurona de entrada a cada neurona intermedia y de cada neurona intermedia a cada neurona en la capa de salida. Cada conexión tiene una ponderación o peso, el cual es almacenado y retenido por la neurona al final de la recepción de la conexión.

La red neuronal opera en dos modos: mapeo y ajuste [SM193]. En el modo de mapeo, la información fluye a través de la red, de las entradas a las salidas. En el modo de ajuste, la información fluye alternadamente hacia delante y hacia atrás.

En el modo de mapeo, la red procesa una muestra (un patrón) cada vez, produciendo un estimado de los valores de las variables dependientes basándose en los valores de las variables independientes de cada muestra. Primero un conjunto de valores de las variables independientes es cargado en la capa de neuronas de entrada de la red. Las neuronas de entrada no realizan un cálculo, cada neurona de entrada sencillamente envía una copia de su valor a todas las neuronas de la capa intermedia. Cada neurona intermedia calcula la suma ponderada de las entradas utilizando sus únicas conexiones de ponderación como de peso, incluyendo en la suma un término de umbral o peso de tendencia.

Enseguida cada neurona intermedia calcula una función sigmoide de su suma. La función sigmoide simplemente acota el valor por debajo de un rango límite. Después de acotar su suma ponderada, cada neurona intermedia envía su resultado a todas las neuronas de salida. Cada neurona de salida desarrolla un cálculo similar. Utiliza pesos almacenados en memoria para calcular la suma ponderada de los valores recibidos de las neuronas intermedias (más un peso de tendencia), entonces calcula una función sigmoide de esta suma. Los valores resultantes de estas salidas son un estimado de lo que representan las variables independientes.

La naturaleza del mapeo llevado a cabo por la red depende de los valores de los pesos. Retropropación es el método de búsqueda de los valores óptimos de las conexiones de ponderación o pesos. Involucra ajustar a la red en muestras para las cuales los valores correctos de las salidas son conocidos.

El ajuste comienza con valores arbitrarios para los pesos, alatorios, y se procesan iterativamente. Cada iteración es llamada un *epoch*. En cada epoch la red ajusta los pesos en la dirección que se reduce el error (la diferencia entre salidas actuales y salidas deseadas o targets). Como el proceso es iterativo de ajustes continuos, los pesos gradualmente convergen a un conjunto óptimo de valores. Muchos epochs son usualmente requeridos antes de que el ajuste se complete.

Para hacer que los pesos se ajusten en cada epoch, la red debe procesar toda la muestra. Por cada muestra en turno, la red lleva a cabo el siguiente proceso de cálculo:

- Primero, la red mapea las entradas para la muestra actual en salidas, usando valores aleatorios para los pesos actuales. En los epochs iniciales las salidas estarán muy alejadas de la exactitud, debido a que los pesos iniciales están lejanos de los correctos.
- Enseñada, las neuronas de salida son informadas de los valores deseados (*targets*) para esa muestra, basándose en la diferencia entre la salida actual y la salida deseada, cada neurona de salida determina la dirección en la cual sus pesos tienen que ajustarse para reducir el error, tanto como la cantidad de cambio pueda ser hecha.
- Cada neurona de salida entonces propaga la cantidad de su error en la muestra actual hacia atrás a través de las conexiones de las neuronas intermedias. Las neuronas intermedias usan esos errores para determinar en qué dirección y cantidad deben cambiar sus pesos, justo cuando las neuronas de salida lo hayan realizado. En el caso de las neuronas ocultas, esos ajustes están diseñados para reducir el error medio a través de todas las salidas.

Por cada muestra se da entonces un paso hacia adelante a través de la red para determinar las salidas actuales de la red, seguido por un paso hacia atrás para determinar, basándose en la diferencia entre esas salidas y las correctas, cómo los pesos deben ser ajustados.

Como las muestras son procesadas, cada neurona de salida y cada neurona intermedia acumulan el cambio deseado en la red en cada uno de sus pesos. Al final del epoch, todas las neuronas intermedias y las neuronas de salida cambian su peso mediante la cantidad apropiada. El proceso es cíclico a través de toda la muestra, acumulando los cambios deseados en cada peso, y realizando los cambios, iterativamente hasta que se alcanzan los valores óptimos.

La clave dentro de Retropropagación es que las neuronas ocultas son capaces de determinar qué tanto cambiar sus pesos si reciben información del error de cada neurona de salida.

El método para calcular el cambio de pesos es el gradiente descendiente. El gradiente descendiente es una técnica familiar en estadística y backpropagation puede ser

interpretado como un método de modelado estadístico. El mapeo realizado durante el paso hacia delante de la red corresponde a la evaluación de una ecuación en la cual las variables independientes son las entradas de la red y las variables dependientes son las salidas de la red. Los valores calculados por las neuronas intermedias son los términos de esta ecuación, los pesos en la red son los coeficientes de la ecuación. El método del gradiente descendiente para ajustar esos coeficientes está diseñado para producir el conjunto de coeficientes que minimiza el error del modelo. La medida del error para backpropagation es el error cuadrático medio, exactamente la misma medida usada en métodos de modelado convencional.

Como podemos ver el entrenamiento mediante la técnica clásica, es complejo y muy dependiente de la obtención del gradiente descendiente. Esto únicamente es para entrenar. El diseño de la red neuronal no puede ser realizado por el algoritmo de Retropropagación mismo. Existe una fuerte distinción entre entrenamiento y diseño de CNAs. Nuestro enfoque, como veremos más adelante, es integral proponiendo una metodología de desarrollo donde simultáneamente se logre diseñar y entrenar un CNA, robusto, sin tendencia de clasificación a una clase determinada, y que los porcentajes de clasificación sean superiores a porcentajes definidos como aceptables

#### 1.4 COMPUTACIÓN EVOLUTIVA

Parte importante de la llamada Computación Adaptable es la Computación Evolutiva (CE), la cual es la metodología de programación que integra las técnicas de los Algoritmos Genéticos, Estrategias Evolutivas, Programación Evolutiva, los métodos basados en el gradiente descendente, y la Programación Genética. La técnica evolutiva que ha ganado mayor popularidad son los Algoritmos Genéticos, los cuales esencialmente son algoritmos de búsqueda basados en mecanismos simplistas, que hacen analogías de la selección natural de las especies y la genética. Combinan sobrevivencia del individuo más apto con un estructurado y a la vez aleatorio intercambio de información entre individuos para construir un algoritmo de búsqueda con algunas de las innovaciones instintivas y heurísticas de la búsqueda humana. En toda generación, un nuevo conjunto de criaturas artificiales (representaciones cromosómicas de redes neuronales) es creado, usando estructuras de datos complejas no uniformes, combinándolas con piezas de las criaturas artificiales más

aptas de generaciones anteriores y ocasionalmente una parte de la nueva generación es más apta que la actual. En primera instancia, los Algoritmos Genéticos no son una simple búsqueda aleatoria. Aprovechan eficientemente información histórica para especular en la búsqueda de nuevos individuos con la esperanza de mejorar el desempeño de los individuos cromosómicos artificiales.

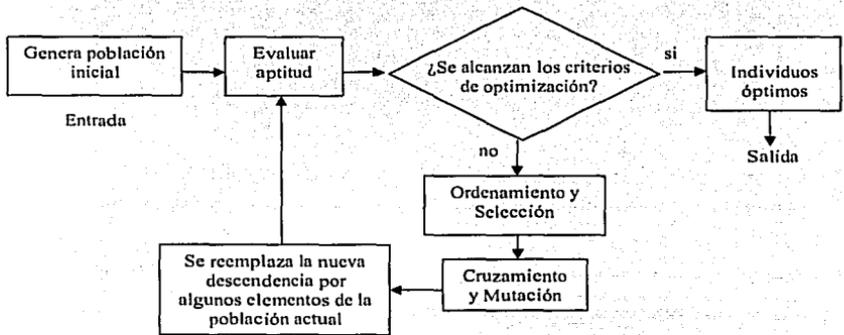


Figura 1.8. Estructura general del Algoritmo Genético

Los Algoritmos Genéticos (AG) son diferentes de la optimización y los procedimientos de búsqueda clásicos en cuatro aspectos:

- Los AG trabajan con un conjunto de parámetros codificado (genotipos) no con los parámetros mismos.
- Los AG buscan dentro de una población de puntos, no en un solo punto.
- Los AG utilizan recompensas (función objetivo) a la información fenotípica, no derivaciones o algún otro tipo de conocimiento.
- Los AG utilizan reglas de transición probabilística, no reglas deterministas.

De forma general, la figura 1.8 ilustra el desarrollo y estructura de un algoritmo genético. En resumen el AG es una técnica de búsqueda aleatoria dirigida, la cual encuentra la solución óptima global dentro de espacios de búsqueda complejos y multidimensionales. Un AG está modelado mediante una analogía con el proceso de selección natural de las especies, los operadores utilizados están inspirados en los procesos biológicos que se dan en las especies de seres vivos. Estos operadores conocidos como operadores genéticos manipulan individuos (posibles soluciones a un problema determinado) sobre varias generaciones para mejorar su desempeño (mejorar la aproximación a la solución óptima) gradualmente.

### 1.5 OPTIMIZACIÓN MULTIOBJETIVO

En los métodos clásicos de entrenamiento neuronal, la obtención del gradiente descendente nos indica la dirección y magnitud de cómo la red neuronal debe ajustar sus parámetros y el error cuadrático medio es la expresión matemática que nos dice que tan bien o mal va el entrenamiento. Incluso usualmente el entrenamiento de una red neuronal se detiene cuando el error cuadrático promedio está por debajo de un valor de sensibilidad previamente definido. La desventaja más obvia del error cuadrático promedio es que siendo una expresión puramente matemática, ésta tiene poca información con respecto a la tarea desarrollada por la red neuronal. El error cuadrático promedio, al ser un indicador exclusivamente cuantitativo, nos oculta información del desempeño de la red, no permite percibir errores cualitativos menores y mayores. En específico, si el objetivo de la red neuronal es ser un CNA, el error cuadrático medio nos dice muy poco a cerca de la frecuencia de errores de clasificación, ya que hace a un lado la naturaleza de los errores potenciales de clasificación.

Nosotros de manera particular queremos entrenar y diseñar CNAs para clasificar patrones de estados de desarrollo celular que pertenecen a tres clases distintas: núcleos, mitosis y basura. Entonces, estamos en un problema multiobjetivo. En optimización multiobjetivo o multicriterio, la noción de optimización no es del todo obvia. Si rechazamos de antemano independizar los valores relativos de los diferentes criterios, debemos proponer una definición diferente de optimización, una que respete la integridad de cada uno de los

critérios de clasificación por separado. Para tal fin vamos a comenzar por enunciar un concepto básico en optimización multiobjetivo, el cual se llama Óptimo de Pareto y vamos a necesitar dos definiciones importantes para plantear de manera formal tal concepto [COE99], [FON95]:

**Definición 1.1 (Dominancia de Pareto en un problema de minimización)** Un vector  $\bar{u} = (u_1, u_2, \dots, u_n)$  se dice domina a  $\bar{v} = (v_1, v_2, \dots, v_n) \Leftrightarrow \bar{u}$  es parcialmente menor que  $\bar{v}$  ( $\bar{u} \prec_p \bar{v}$ ), es decir,

$$\forall i \in \{1, 2, \dots, n\}, u_i \leq v_i \wedge \exists i \in \{1, 2, \dots, n\} \text{ tq } u_i < v_i$$

**Definición 1.2 (Óptimo de Pareto)** Una solución  $x_p \in U$  se dice óptimo de Pareto si y solo si no hay  $x_p \in U$  para el cual  $\bar{v} = f(x_p) = (v_1, v_2, \dots, v_n)$  domina a  $\bar{u} = f(x_p) = (u_1, u_2, \dots, u_n)$ .

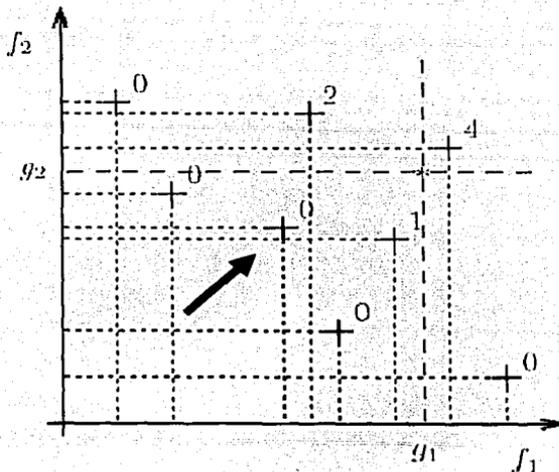


Figura 1.9 Ordenamiento de una función con dos objetivos, mediante la técnica de Pareto

En esencia Pareto asigna el “*ranking*”, jerarquía o dominancia de un individuo dentro de la población considerando el número de individuos que lo dominan en la actual población. Los individuos no dominados (los dominantes) tienen ranqueo 0, mientras que los dominados son penalizados de acuerdo a la concentración de la población en el conjunto de individuos no dominados. Con las definiciones anteriores, analicemos la figura 1.9. En la gráfica bidimensional se muestran los puntos de una función con dos objetivos que se desea minimizar, ambos objetivos tienen la misma importancia. El eje de las abscisas representa los valores correspondientes al criterio uno,  $f_1$ , y el eje de las ordenadas son los valores correspondientes al criterio dos,  $f_2$ . Los pares ordenados marcados con valor cero corresponden al conjunto de óptimo de Pareto; esto es, son los elementos no dominados, los dominantes de esa muestra de puntos. Ese conjunto de óptimo de Pareto aparentan ser las mejores posibilidades de elección, aunque ninguno de los cinco puntos es el mejor en ambas direcciones. Podemos ver que hay alternancia en la ganancia de uno a otro de esos cinco puntos; hay ganancia a lo largo de una dimensión y pérdida en la otra. La solución, señalada con una flecha en la gráfica de la figura 1.9, es mediatizada y recibe esa denominación debido a que es un individuo con un desempeño aceptable, pero no sobresaliente en ninguna dimensión en particular, quizá con un desempeño promedio en todas las dimensiones de la función objetivo [COE99]. De forma extrema el punto marcado con 4 parece ser la elección más débil, esto es porque es dominado por cuatro elementos de la muestra. Como práctica común, el concepto de optimización de Pareto no nos ayuda a seleccionar una sola alternativa del conjunto de no dominados. La decisión debe hacerla finalmente un juicio de ponderación entre las alternativas para llegar a una decisión particular.

Como podemos ver de la figura 1.9, la optimización multiobjetivo, no nos va ofrecer una solución única, sino un conjunto óptimo de Pareto. Esto es, un conjunto de soluciones que de acuerdo a los criterios de optimización que hayamos definido, ninguna es mejor que otra y son las mejores soluciones que cumplen simultáneamente todos los criterios.

En la figura 1.9 los puntos (las soluciones) pertenecen a una función con dos objetivos. Ese tamaño de dimensión nos permite visualizar con cierta facilidad el conjunto de óptimo de

Pareto, aun si la función se tratara de tres objetivos, la visualización del conjunto de óptimo de Pareto se podría distinguir, claro con un poco más de dificultad, pero solo se trataría de ajustar la perspectiva de la gráfica para tener una idea de la zona donde se encuentran los elementos del óptimo de Pareto. Pero si el número de criterios a optimizar se incrementa a cuatro o más dimensiones, ya no es tan fácil, más bien imposible visualizar gráficas en más de tres dimensiones. Para tal situación vamos a utilizar un artificio para simplificar la visualización de funciones con múltiples objetivos. Este método se llama Método de Coordenadas Paralelas[ROD99]. Este consiste en asociar un índice  $i$  a cada objetivo y representar un punto no dominado por medio de una línea que une los puntos  $(i, f_i^*(\bar{x}))$ , donde  $f_i^*$  es la normalización en el intervalo  $[0,1]$  del  $i$ -ésimo criterio de la función multicriterio  $f$  evaluada en el vector  $\bar{x}$ . En la figura 1.10 ponemos un ejemplo de este artificio de visualización multiobjetivo.

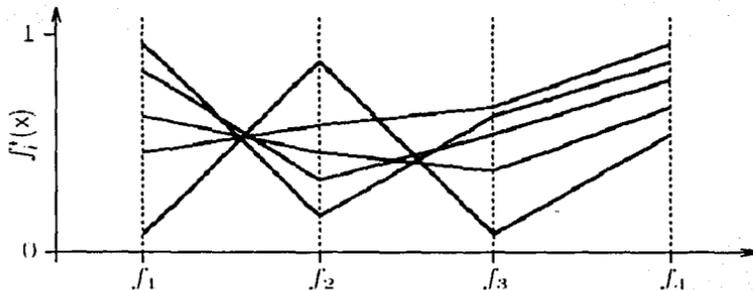


Figura 1.10. Visualización del conjunto de elementos no dominados de una Función de cuatro objetivos, mediante el Método de Coordenadas Paralelas.

En la figura 1.10 mostramos un ejemplo de un conjunto de cinco elementos no dominados para una función con cuatro objetivos. Al igual que en la figura 1.9, los elementos dominantes de este ejemplo presentan alternancias en las ganancias entre los distintos

objetivos de la función. Con el Método de Coordenadas Planas o Paralelas, ya no hablamos de alternancia en la ganancia de una dimensión a otra, sino de un índice que posiciona un criterio de una función a otro índice, es decir, lo que el Método de Coordenadas Paralelas realiza es una proyección de un espacio de  $\mathbb{R}^n$  a  $\mathbb{N}$ , mediante la asignación uno a uno de las dimensiones de una función multiobjetivo con un conjunto de índices en los números naturales. Esta forma de visualización será muy útil para analizar los resultados que alcancemos con este trabajo.

La técnica de Pareto parece ser mejor que la medida del error cuadrático promedio, ya que ordena las soluciones potenciales tomando en cuenta información con varias dimensiones, además de que permite hacer consideraciones simultáneas cualitativas y cuantitativas. Aún así este enfoque presenta inconvenientes. En la muestra de puntos de la gráfica en la figura 1.9, no tenemos información adicional de la jerarquía entre ambos objetivos y todos los puntos del conjunto de óptimo de Pareto tienen el mismo valor de ranqueo u ordenamiento cero; pero esto no quiere decir que genotípicamente y fenotípicamente sean muy semejantes. Así el concepto de Pareto no nos permite distinguir de manera más estricta a los elementos más óptimos, ya que de los elementos del óptimo de Pareto, específicamente los elementos que se encuentran en los extremos laterales del frente de Pareto, algunos pueden pasar por dominantes sin serlo de forma estricta[DEB01]. En nuestra aplicación, un CNA que se sitúe en los extremos laterales del frente de Pareto significa un CNA sesgado. Así, esta forma de asignación de desempeño induce un error cuando no se tiene bien definida la función multiobjetivo, cuando las jerarquías de los varios objetivos no están bien determinadas y cuando las restricciones de la misma función multiobjetivo no se han determinado de forma correcta.

## 1.6 DERIVA GENÉTICA

Como ya mencionamos, la Computación Evolutiva, son artificios inspirados en los procesos de selección y evolución natural. La analogía artificial de fenómenos naturales por lo regular, trae consigo que los fenómenos que se presentan en los modelos naturales originales tengan su contraparte en los modelos artificiales. En la naturaleza, en específico en los modelos poblacionales naturales, se presenta el fenómeno de endogamia, que a nivel

cromosómico lo podríamos ver como la llamada deriva genética (genetic drift)[DEB89]. Endogamia significa engendrar "desde adentro" y en un principio, significa matrimonio entre personas de un mismo linaje. Este fenómeno se observa consistentemente cuando un grupo pequeño de individuos se ve aislado y toda su variedad de selección y desarrollo se ve reducida al mismo grupo de individuos. En los algoritmos evolutivos y genéticos multiobjetivo, este fenómeno también se presenta, ya que funcionan con poblaciones de soluciones potenciales finitas, estáticas, no muy grandes, cuyo intercambio cromosómico se reduce al intercambio entre las características genotípicas con los mismos miembros de la población de soluciones factibles actuales. En el desarrollo de las generaciones de tales algoritmos, el intercambio cromosómico constante con individuos de la misma población, podríamos decir linaje, provoca que la calidad de la dispersión cromosómica se corrompa, de forma que se de la replicación de individuos dominantes y con ello se provoque la convergencia o especialización de una especie hacia un criterio determinado, resultado de la poca variedad en la disponibilidad cromosómica. También sabemos que en los algoritmos evolutivos simples y en los multiobjetivos la herencia de las primeras generaciones tiene un carácter determinante en las generaciones sucesivas. Por ello, siempre se desea que la variedad de la población inicial sea lo más dispersa posible, pero el error estocástico asociado con los operadores genéticos durante las generaciones sucesivas causa que la población converja o se especialice en una u otra función determinada. Este efecto tiende a ser más notorio en problemas multiobjetivo que en problemas uniobjetivo, ya que la replicación cromosómica tiende a favorecer la especialización de los grupos aislados, lo cual no siempre es deseable cuando se quiere que los elementos dominantes se especialicen o distribuyan sus capacidades en más de un objetivo. Aunado a esto, la forma en que Pareto plantea el ordenamiento, en algunas aplicaciones no es muy adecuada, ya que según las características propias de la naturaleza del problema, algunos elementos del frente de Pareto pueden representar soluciones que no son no dominadas. Por otro lado existen varias técnicas que no emplean el concepto de Pareto, como la Articulación de Preferencias a Posteriori, el Ordenamiento Lexicográfico o VEGA, entre otras, todas ellas tienen ventajas y desventajas; entre una de las desventajas importantes es que predeterminan la existencia de una tendencia a eliminar los elementos mediatizados. Este problema de tender contra los puntos medios es serio. Durante una generación dada no debe de haber tendencia

alguna contra los elementos localmente dominados. Si aceptamos el planteamiento de optimización de Pareto, esos individuos y todos los demás deben de tener el mismo potencial reproductivo para asegurar la dispersión genotípica y fenotípica dentro de la población de soluciones potenciales. Mucho se ha realizado en el sentido de combinar técnicas de optimización multiobjetivo basadas en técnicas evolutivas, con técnicas basadas y no basadas en el concepto de Pareto, lo cual ha dado resultados prometedores, pero la mayoría de los resultados sólo incluyen problemas de optimización de funciones analíticas, lo cual de alguna forma condiciona a que el problema puede ser analizado de manera a priori [COE99], [DEB01], [FON95], [ROD99].

Estas situaciones anteriores nos conducen a realizar ciertas modificaciones a las de técnicas no convencionales, modificaciones que nos resuelvan un problema particular, pero que puedan ser útiles para resolver una problemática general: el diseño y entrenamiento automático de CNAs.

### 1.7 ARQUITECTURA BÁSICA DEL CNA

La manera más generalizada de implementar un CNA para clasificar múltiples clases es empleando una neurona de salida separada por cada clase [MAS93]. Así el ajuste de pesos requiere que la neurona correspondiente a la clase determinada se active (se prenda), mientras que el resto de las neuronas de salida permanezcan desactivadas (apagadas). Además cada patrón de entrenamiento tiene asociado un vector deseado, el cual indica, durante la etapa de ajuste, la clasificación a la que pertenece determinado patrón. Este patrón deseado o target es un vector de dimensión tres, dado que se tienen tres clases a distinguir. Esto se refleja en correspondencia con el número de neuronas en la capa de salida del clasificador, así cada neurona de la capa de salida corresponde a una clase de desarrollo de las células de los patrones de entrenamiento. Finalmente, dadas las características de los patrones de entrenamiento, la arquitectura básica del CNA que estamos proponiendo tiene cinco neuronas en la capa de entrada, una sola capa intermedia con tres neuronas y tres neuronas en la capa de salida, como se muestra en la figura 1.7. Como se dijo anteriormente, se cuenta con trabajos previos reportados, de donde se extrajo la arquitectura anterior. Dado que fue la que mejor resultados dio; aprovechamos la

experiencia de tales trabajos y comenzamos por esta misma configuración exitosa. Es esta estructura básica de CNA la que será evolucionada en nuestra propuesta de diseño y entrenamiento de CNAs. Pero no será en el sentido como lo plantean los métodos clásicos neuronales, los cuales proponen iniciar con una red neuronal cuya arquitectura tenga un número considerable de neuronas y entonces comenzar a eliminar neuronas, en base a información muy específica de la influencia de los pesos de ciertas neuronas, hasta alcanzar una red neuronal sin bajo ni sobre ajuste como es el caso de los llamados métodos de Poda Neuronal. Otra opción implica optar por tomar el enfoque complementario, comenzar con una arquitectura neuronal con pocas neuronas e ir anexando nuevas neuronas, cada vez realizando un mapeo de la Red Neuronal para ir ponderando la inclusión de tal nueva neurona, como es el caso de las técnicas de Crecimiento Neuronal. En nuestra propuesta, como veremos más adelante, los mecanismos de poda y crecimiento neuronal, los plantearemos mediante operadores genéticos de cruzamiento y mutación diseñados de acuerdo al tipo de CNA que necesitamos.

## 1.8 RESUMEN

En este capítulo analizamos las características estadísticas de la población de patrones de etapas de desarrollo celular. Ese análisis nos ayudo a tomar en consideración el sesgo implícito en la población de patrones. Con el estudio realizado y las consideraciones necesarias, exploramos las posibilidades de solución al problema del diseño y ajuste de un CNA para clasificar los patrones de etapas de desarrollo celular. Revisamos tres conceptos importantes para tal fin: el Algoritmo de Retropropagación, la Computación Evolutiva y la Optimización Multiobjetivo. Los tres conceptos son muy útiles para saber cual es el estado del arte en lo referente al problema que nos interesa, ya que nos permiten visualizar cual es el camino a seguir, que técnicas son las adecuadas a seguir, y cuales debemos desechar. Pero más importante, nos advierten de cual es la problemática a la que nos estamos enfrentando para el diseño y entrenamiento automático de CNAs para el reconocimiento de patrones de estados celulares.

En el capítulo dos vamos a proponer una metodología integral, para alcanzar los objetivos que nos planteamos en el principio de este trabajo y que solucionen la problemática aquí presentada.

TESIS CON  
FALLA DE ORIGEN

## CAPÍTULO DOS

### METODOLOGÍA INTEGRAL PARA EL DISEÑO Y ENTRENAMIENTO DE CNA's MEDIANTE UN ALGORITMO EVOLUTIVO MULTIOBLACIONAL Y MULTIOBJETIVO

#### 2.1 METODOLOGÍA

En el capítulo uno analizamos la problemática involucrada en el diseño y entrenamiento de CNA's. La cantidad de variables de configuración estructural involucradas tienen estrechas relaciones que entran en conflicto, esto hace que el diseño de una red neuronal se convierta en un problema complejo. En nuestro caso particular deseamos obtener un CNA, robusto que clasifique bien los patrones de una muestra de datos experimental obtenida de imágenes biomédicas y por lo tanto también debemos tener en cuenta las características particulares de la muestra experimental de patrones de estados celulares.

Cuando nos referimos a un CNA robusto nos referimos a que queremos un CNA o un conjunto de CNA's que funcionen bien, no solamente para una determinada muestra de datos experimentales ( con los que fuera entrenado ) sino que tenga buenas probabilidades que con una muestra de datos nueva funcione de igual manera. En la figura 1.1 observamos que el componente final de un sistema reconocedor de patrones es un elemento llamado clasificador. El problema de clasificación más simple es el binario. En un problema así, el clasificador realiza una decisión positiva o negativa al respecto de un elemento en particular que se desea clasificar. Pero la mayoría de las veces decidir positiva o negativamente no es lo más conveniente, sobre todo en los problemas reales donde la mayoría de los problemas de clasificación involucran clasificar elementos que potencialmente pertenecen a más de dos clases. Incluso, algunas veces dentro de la problemática de clasificación además de la multclasificación, va incluida la ambigüedad, lo cual significa que las clases o conjuntos de elementos pueden ser no disjuntos y un elemento puede pertenecer a dos o más clases simultáneamente. Como dijimos en la introducción, existen dos tipos principales de algoritmos para entrenar reconocedores de patrones. Nosotros específicamente vamos a diseñar y emplear un clasificador estadístico, en particular un clasificador neuronal artificial (CNA). Las redes neuronales son utilizadas para resolver diferentes tipos de problemas, uno de los más comunes es para resolver

problemas de clasificación. A este respecto existe toda una línea de investigación la cual incluye: análisis de datos, técnicas de diseño, metodologías de entrenamiento, estrategias de toma de decisiones y validación de resultados. La conjunción de estas actividades complementan las tareas del experto en redes neuronales, quien desarrolla tales tareas según su conocimiento y experiencia, regularmente mediante un proceso de prueba y error. Aunado a esto, el experto generalmente tiende a desarrollar separadamente tales actividades siendo que algunas de ellas deben ser desarrolladas simultáneamente para obtener mejores resultados. También es práctica usual particularizar la solución de acuerdo a las características singulares del problema a resolver.

Una de nuestras metas es no detenernos en dar una solución particular para un problema. Dentro de nuestros objetivos principales está ofrecer una metodología y consecuentemente una herramienta, que nos proporcione un conjunto de soluciones para resolver toda la problemática involucrada y específicamente, como caso de prueba, proporcionamos una solución robusta al problema de clasificación de patrones de estados de desarrollo celular. Así, pensamos que nuestro enfoque se aproxima más a una metodología que a una técnica ya que la diferencia entre método y técnica está en que la técnica representa las etapas operacionales limitadas, unidas a elementos prácticos, concretos adaptados a un fin definido, mientras que el método es una concepción intelectual que coordina un conjunto de operaciones y en general, diversas técnicas. Con estos fines vamos a describir el conjunto de técnicas que van a conformar nuestra propuesta de metodología para el diseño y entrenamiento de CNAs.

## 2.2 MUESTREO Y REPRESENTATIVIDAD

La manera de recolectar los datos es por lo menos tan importante como la forma de analizarlos. En particular, una muestra debe ser representativa de la población, y a menudo el muestreo aleatorio es la mejor manera para lograr representatividad [WON98]. Si una muestra no es aleatoria, puede estar tan sesgada que es más que inútil. Una muestra se denomina muestra aleatoria simple si cada individuo de la población tiene la misma probabilidad de ser elegido cada que se extrae una observación. La aleatoriedad incrementa la probabilidad de que una muestra sea representativa de la población donde se extrajo tal muestra. Por supuesto, el objetivo del muestreo aleatorio es efectuar una inferencia acerca

de la población; así entonces, esperamos que la media muestral  $\bar{X}$  sea una estimación aproximada de la media de la población  $\mu$ . Es probable que una media muestral se encuentre un poco por arriba o por debajo de la media de la población. El valor que indica que tan arriba o abajo es  $\bar{X}$  una buena estimación de  $\mu$  es el error estándar  $EE$ , el cual se determina con ayuda de  $\sigma^2$ , la varianza de la población, mediante la siguiente formula:

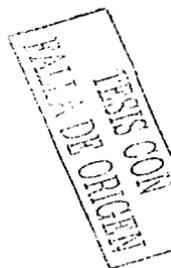
$$EE = \frac{\sigma}{\sqrt{n}}$$

Ecu. 2.1

Esta fórmula muestra que mientras más grande el valor de  $n$ , que representa el tamaño de la muestra, más pequeño se hace el error estándar; es decir, a medida que crece el tamaño  $n$  de la muestra el error estándar decrementa. Formalmente, mientras más grande sea la muestra  $X$ , con mayor exactitud se estima la media de la población. De hecho, la cantidad sola puede ser bastante engañosa. Una muestra grande pero sesgada puede ser aceptable debido a su tamaño, aunque de hecho solamente consiste en el mismo sesgo repetido una y otra vez, es decir, la calidad es más importante que la cantidad.

Clase	Frecuencia	Frecuencia relativa $p(x)$ a la población	Cálculo de la media $\mu$ de cada parámetro		Cálculo de la varianza $\sigma^2$
Basura	387	0.4300	Forma	0.7396	0.0255
			Área	0.1908	0.0214
			N. gris	0.6400	0.0106
			Dep. hoyos	0.3352	0.0090
			D.E.N.G.	0.6791	0.0183
Mitosis	182	0.2022	Forma	0.7514	0.0316
			Área	0.3392	0.0381
			N. gris	0.7446	0.0088
			Dep. hoyos	0.2502	0.0071
			D.E.N.G.	0.6363	0.0134
Núcleos	331	0.3678	Forma	0.9174	0.0025
			Área	0.0718	0.0007
			N. gris	0.5065	0.0053
			Dep. hoyos	0.4624	0.0082
			D.E.N.G.	0.4684	0.0141
Total	900	1			

Tabla 2.1 Valores estadísticos de la población total de los patrones de estados celulares.



En la tabla 2.1 se muestran los valores de las distribuciones de frecuencia de la población de patrones de etapas de desarrollo celular, así como los valores de las medias y las varianzas de cada uno de los cinco parámetros ( forma, área, N. gris, Dep. hoyos, Desviación Estándar del Nivel de Gris DENG ) que caracterizan a los tres estados de desarrollo celular: mitosis, núcleos y basura.

Es obvio que existe una disparidad proporcional implícita en la naturaleza de la población experimental de patrones de estados celulares. De la tabla 2.1 podemos ver que las proporciones de patrones correspondientes a basura, mitosis y núcleos son aproximadamente del 46%, 20% y 37% respectivamente. Luego entonces, en la población tenemos más del doble de patrones correspondientes a basura que los correspondientes a mitosis y poco menos de dos veces el número de patrones de núcleos que de mitosis. En conclusión, esta diferencia proporcional se acentúa en la clase mitosis, así que representativamente la clase mitosis frente a las clases núcleos y basura es débil. La proporción, como ya se dijo en el capítulo uno, puede ser un problema durante el entrenamiento del CNA mediante una técnica clásica neuronal. Pero como nos refieren los expertos biomédicos, la disparidad proporcional es propiedad de los estados de desarrollo de las poblaciones celulares existentes en la naturaleza. Entonces nosotros en la intención de que la metodología sea lo suficientemente adaptable, autónoma e independiente del problema que se quiera resolver, debemos respetar las características del problema y proponer que la metodología de solución sea la que se adapte al problema y no suceda lo contrario. Así las proporciones de las muestras seleccionadas de entrenamiento no deben inducir ni influir en alguna medida el sesgo natural que produce la representatividad proporcional de una clase determinada en un CNA. Estadísticamente podemos denominar que una muestra es típica o representativa de la población total, si los valores de las medias muestrales de los parámetros que caracterizan a cada clase de la población experimental, están dentro del intervalo establecido por la medida de la media de cada parámetro de la población, más menos el error estándar descrito en la ecuación 2.1 [WON98]. Esta medida de acotación nos define los intervalos típicos muestrales de la población total para cada parámetro. En las tablas 2.2 a 2.4 se muestran los intervalos típicos muestrales para cada clase de patrones respecto de la población total de patrones de etapas de desarrollo celular.

Intervalo típico Forma	Intervalo típico Área	Intervalo típico N. gris	Intervalo típico Dep. hoyos	Intervalo típico D.E.N.G.
(0.7383, 0.7646)	(0.3248, 0.3537)	(0.7377, 0.7516)	(0.2440, 0.2565)	(0.6277, 0.6449)

Tabla 2.2 Intervalos típicos muestrales para la clase mitosis.

Intervalo típico Forma	Intervalo típico Área	Intervalo típico N. gris	Intervalo típico Dep. hoyos	Intervalo típico D.E.N.G.
(0.9147, 0.9202)	(0.0704, 0.0733)	(0.5025, 0.5105)	(0.4574, 0.4674)	(0.4619, 0.4750)

Tabla 2.3 Intervalos típicos muestrales para la clase núcleos.

Intervalo típico Forma	Intervalo típico Área	Intervalo típico N. gris	Intervalo típico Dep. hoyos	Intervalo típico D.E.N.G.
(0.7315, 0.7478)	(0.1833, 0.1982)	(0.6348, 0.6453)	(0.3304, 0.3401)	(0.6722, 0.6860)

Tabla 2.4 Intervalos típicos muestrales para la clase basura.

En redes neuronales usualmente se emplea la validación cruzada (*cross-validation*)[HAY98], la cual es una herramienta estadística que particiona aleatoriamente el conjunto de datos experimentales en una muestra de entrenamiento y una muestra de validación y posteriormente la muestra de entrenamiento es nuevamente particionada en dos subconjuntos disjuntos: un subconjunto de estimación, el cual es empleado para seleccionar la estructura de red neuronal con los mejores pesos; y el subconjunto de validación se usa para probar o validar la red neuronal encontrada con el subconjunto de estimación. El conjunto o muestra de validación se utiliza para medir el desempeño de generalización de la red neuronal. Existen algunas variantes de la validación cruzada, las cuales toman en cuenta la escasez de cardinalidad de la muestra experimental de datos. Como por ejemplo *multifold cross-validation* y el *leave-one-out*. En esencia tales modificaciones dividen el conjunto de datos disponibles en  $K$  subconjuntos, con  $K > 1$ . Así la red neuronal es entrenada alternadamente con  $K > 1$  subconjuntos y la validación se realiza con el subconjunto restante. Tales enfoques de selección de muestras de entrenamiento y validación apuestan su capacidad de representatividad y consecuentemente de generalización, completamente en la aleatoriedad en la selección de la partición de los subconjuntos. Pero como dijimos anteriormente, una muestra grande o varias pequeñas,

# TESIS CON FALLA DE ORIGEN

## CAPÍTULO DOS

según el caso pueden ser aceptables pero también estar sesgadas, hecho que solamente consiste en el mismo sesgo repetido una y otra vez; es decir, la calidad es más importante que la cantidad. Así el sesgo proporcional, propio de las poblaciones experimentales de etapas de desarrollo celular, lo estamos mitigando mediante la selección uniformemente aleatoria condicionada por los intervalos típicos muestrales por parámetros de los patrones de cada clase, como lo mostramos en las tablas 2.5 y 2.6.

Clase	Frecuencia	Frecuencia relativa $p(x)$ a la muestra	Frecuencia relativa $p(x)$ a la población	Cálculo de la media $\bar{X}$ de cada parámetro	
Basura	116	0.4028	0.1289	Forma	0.7449
				Área	0.1862
				N. gris	0.6351
				Dep. hoyos	0.3344
				D.E.N.G.	0.6771
Mitosis	73	0.2535	0.0811	Forma	0.7611
				Área	0.3394
				N. gris	0.7462
				Dep. hoyos	0.2522
				D.E.N.G.	0.6326
Núcleos	99	0.3438	0.1100	Forma	0.9196
				Área	0.0716
				N. gris	0.5095
				Dep. hoyos	0.4638
				D.E.N.G.	0.4653
Total	288	1	0.3200		

Tabla 2.5 Valores estadísticos de cada clase de la muestra de entrenamiento A.

La metodología que estamos proponiendo selecciona el conjunto de patrones de entrenamiento, de manera que sea representativo de la población total de los patrones de estados celulares. Nuestro enfoque toma la práctica común en el entrenamiento de redes neuronales, particionar la población total de patrones experimentales en una muestra de entrenamiento y una muestra de validación. La muestra de entrenamiento debe ser menor en tamaño que la de validación, pero sobre todo, la muestra de entrenamiento debe ser representativa de la población de datos experimentales. Esta representatividad la estamos

planteando mediante la contención en los intervalos típicos muestrales. En las tablas 2.5 y 2.6 se muestran dos ejemplos distintos de muestras de entrenamiento, pero con la misma capacidad de representatividad, medida de acuerdo al grado de contención en los intervalos típicos muestrales definidos en las tablas 2.2 a 2.4. Observemos con detenimiento la tabla 2.5, podemos decir que se trata de una muestra típica ya que los valores de la media de cada parámetro están contenidos en los intervalos típicos muestrales mostrados en las tablas 2.2 a 2.4. Obsérvese que la muestra de entrenamiento A de la tabla 2.5 representa aproximadamente el 32% de la población de patrones y que las proporciones entre las clases de patrones no son homogéneas, ya que tenemos aproximadamente, respecto a la población experimental de patrones, 13%, 8% y 11% y respecto a la muestra aproximadamente, 40%, 25% y 34% para Basura, Mitosis y Núcleos, respectivamente, en ambos casos. Dados los valores de las frecuencias relativas a la muestra y a la población, no obstante la muestra A de la tabla 2.5 es representativa de la población de patrones de desarrollo celular.

Ahora veamos en la tabla 2.6 la muestra de entrenamiento B. En ella, a diferencia de la muestra de entrenamiento A, las proporciones entre las tres clases son homogéneas tanto relativamente a la muestra como relativamente a la población aproximadamente 33% y 6%, respectivamente. Aunque la muestra de entrenamiento A es muy distinta en proporciones a la muestra de entrenamiento B, ambas muestras son igualmente representativas de la población experimental de patrones de estados de desarrollo celular, ya que también los valores de las medias de cada parámetro de las tres clases, en la muestra de entrenamiento B, están dentro de los intervalos típicos muestrales.

Ambas muestras de entrenamiento de los ejemplos A y B, mostradas en las tablas 2.5 y 2.6 respectivamente, fueron seleccionadas mediante una distribución uniformemente aleatoria, cuidando siempre que tales muestras de entrenamiento tengan un grado de representatividad igual a cinco[MAT92]. El grado de representatividad lo determinamos mediante la comprobación de que los valores de la media de cada parámetro, de cada clase de la muestra de entrenamiento, estén dentro del intervalo típico muestral correspondiente. Nótese que de esta manera estamos asegurando la representatividad, no únicamente en términos de la selección aleatoria de la muestra de entrenamiento, sino también en el grado de representatividad tomando en cuenta el número de parámetros de los patrones de la

población experimental. De esta forma, la muestra de entrenamiento no únicamente es representativa de la población general sino también es representativa en forma particular de todas las clases que forman a la población.

Clase	Frecuencia	Frecuencia relativa $p(x)$ a la muestra	Frecuencia relativa $p(x)$ a la población	Cálculo de la media $\bar{X}$ de cada parámetro	
Basura	55	0.3333	0.0611	Forma	0.7324
				Área	0.1871
				N. gris	0.6383
				Dep. hoyos	0.3311
				D.E.N.G.	0.6731
Mitosis	55	0.3333	0.0611	Forma	0.7526
				Área	0.3357
				N. gris	0.7419
				Dep. hoyos	0.2479
				D.E.N.G.	0.6352
Núcleos	55	0.3333	0.0611	Forma	0.9165
				Área	0.0708
				N. gris	0.5096
				Dep. hoyos	0.4583
				D.E.N.G.	0.4632
<b>Total</b>	<b>165</b>	<b>1</b>	<b>0.1833</b>		

Tabla 2.6 Valores estadísticos de cada clase de la muestra de entrenamiento B.

### 2.3 DESCRIPCIÓN DEL ALGORITMO EVOLUTIVO MULTIPOBLACIONAL.

Nuestra propuesta de desarrollo y entrenamiento de un CNA esta basada en un enfoque de las técnicas evolutivas. Como se mencionó en la introducción, hay muchos trabajos en este sentido cuyas técnicas de programación obviamente son no convencionales y pertenecen a la llamada Computación Adaptable[YAO99]. Desde este enfoque podemos decir que, dado que estamos estableciendo una metodología de diseño y entrenamiento automático de CNAs, y consecuentemente para comprobar nuestras afirmaciones, estamos desarrollando la herramienta de programación necesaria para hacer evolucionar los CNAs, y que un CNA es también una herramienta de programación para realizar un proceso específico de clasificación. Entonces, lo que estamos haciendo de forma práctica lo podemos ver como un programa de propósito general que diseña y evalúa programas con una funcionalidad

específica basada en una heurística predeterminada. Nuestra heurística o estrategia global de desarrollo y entrenamiento de CNAs esta esquematizada en la figura 2.1

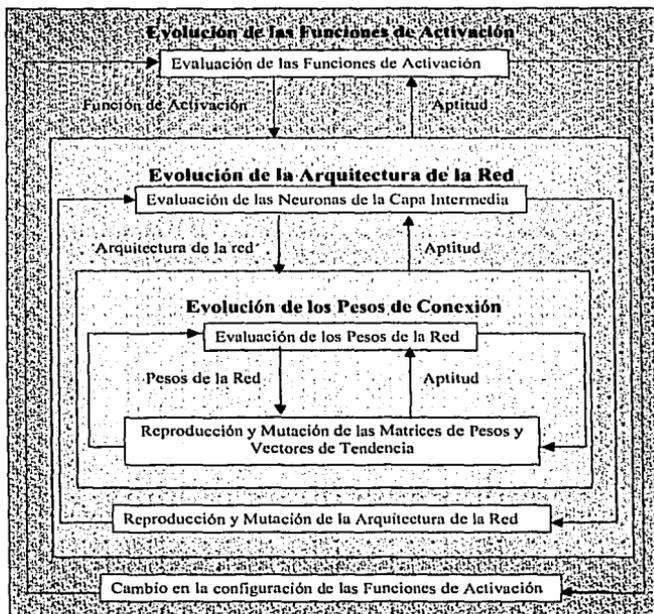


Figura 2.1 Esquema de la Estrategia General para el Desarrollo y Entrenamiento de los CNAs

La figura 2.1 refleja una estrategia no convencional. No estamos proponiendo una estrategia secuencial. Si bien cada módulo de evolución está bien definido, las interacciones entre módulos no son secuenciales, no se está esquematizando un diagrama de flujo convencional con una secuencia de flujo lineal. Lo que estamos proponiendo es un enfoque evolucionista en todos los sentidos, cuyas etapas de desarrollo y entrenamiento están traslapadas, donde el paralelismo y la simultaneidad están implícitas. En la figura 2.1 podemos ver que los bloques de desarrollo y entrenamiento de los CNAs están incrustados dentro de otros bloques de desarrollo y entrenamiento: el primer bloque es de evolución de

los pesos de conexión, el segundo es de evolución de la arquitectura de los CNAs y el tercero es de evolución de las funciones de activación. Incluso el orden lo podríamos invertir, pero la certeza sería que los tres procesos de evolución se están llevando a cabo en forma simultánea. En nuestra propuesta de metodología los CNAs nunca dejan de evolucionar los pesos de conexión ni las funciones de activación. Si bien, más adelante hacemos una distinción en cuanto a desarrollo y entrenamiento de CNAs con estructura fija y con plasticidad en la capa intermedia, es solo para distinguir de alguna forma como se va dando el desarrollo evolutivo de los CNAs en nuestra propuesta de metodología, pero en realidad el enfoque evolucionista que estamos dando es completamente simultáneo.

### 2.3.1 DESCRIPCIÓN DE LA ESTRUCTURA CROMOSÓMICA NO HOMOGÉNEA

Para ir refinando la definición de nuestra metodología, comencemos por describir el cromosoma que empleamos. Se trata de una séxtupla  $(W1, B1, f_1, W2, B2, f_2)$ , tal que:  $W1$  y  $W2$  son las matrices de pesos cuyos elementos  $w_{i,j}^k \in \mathfrak{R}$ , con  $k = 1,2$  significan el peso que va de la  $i$ -ésima neurona a la  $j$ -ésima neurona en la  $k$ -ésima capa de la red,  $B1$  y  $B2$  son los vectores que contiene los términos de tendencia  $b_i^k \in \mathfrak{R}$  o bias de la  $i$ -ésima neurona en la  $k$ -ésima capa, y  $f_k$  es la función de activación de la  $k$ -ésima capa de la red, tal que la imagen de  $f_k$ , a excepción de la función lineal, esta acotada en los intervalos  $[-1,1]$  o  $[0,1]$  y es una función derivable. El conjunto  $FActiv$  de funciones de activación que utilizamos se muestra en la figura 2.2.

Con el fin de obtener mejores resultados de clasificación: los valores de las funciones de activación del CNA, los parámetros de los patrones de entrenamiento y los valores de los vectores deseados se deben encontrar en las mismas magnitudes. Como mencionamos en el capítulo uno, fue necesario realizar un preprocesamiento a los patrones de estados de desarrollo celular, lo que implicó normalizar todos los valores de los parámetros en el intervalo  $[0,1]$ , y los vectores de salida deseada para la clase mitosis, núcleos y basura los definimos como  $[1,0,0]$ ,  $[0,1,0]$  y  $[0,0,1]$ , respectivamente, ya que se quiere clasificar más de dos clases y por que optamos por activar una neurona de salida por cada clase. Así para los propósitos del diseño automático de CNAs, todas las funciones del conjunto  $FActiv$  en la figura 2.2, funcionan igualmente bien y muy convenientemente con nuestro enfoque

evolucionista, las funciones pueden ser intercambiables. Los valores de los elementos en los vectores deseados y las diferencias en las cotas de los intervalos de las imágenes de las funciones de activación, pueden establecer una preferencia por una u otra función de activación, debido a que las respuestas de las neuronas de salida de los CNAs deben caer en esos valores acotados. Sin embargo, tales cotas no son determinantes para establecer los posibles rangos de valores de los patrones de etapas de desarrollo celular, es por esto que también incluimos a la función lineal. Mas como se mencionó los valores de los parámetros de los patrones se normalizaron, por lo tanto la gran mayoría de las funciones de *FActiv* están acotadas en los intervalos  $[-1,1]$  o  $[0,1]$ .

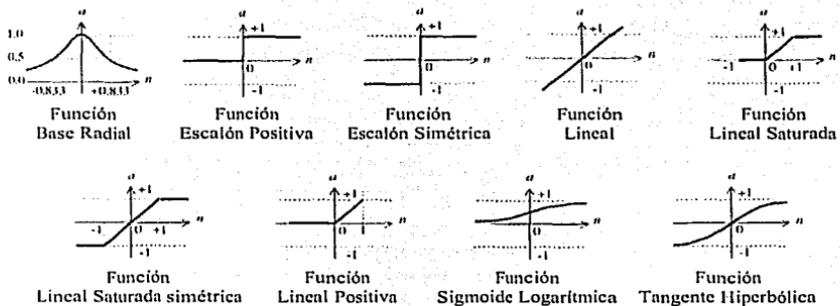


Figura 2.2 Conjunto *FActiv* de funciones de activación para las configuraciones de los CNAs en el desarrollo de la Metodología Evolutiva

El hecho de utilizar funciones derivables obedece a que también se consideró utilizar la metodología evolucionista de diseño y entrenamiento de CNAs como una primera aproximación para encontrar las condiciones iniciales de configuración y entrenamiento, y posteriormente mediante alguna técnica neuronal clásica, refinar el entrenamiento del CNA hasta alcanzar la aproximación que se considere suficiente a los resultados deseados.

En el capítulo uno discutimos la problemática inmersa en el diseño y entrenamiento de CNAs. También analizamos las características propias del problema de clasificación de patrones de estados celulares, inclusive en la introducción discutimos los trabajos que en

cierta forma son base de nuestra propuesta. Con toda esta experiencia previa comenzamos determinando una población inicial de CNAs con estructura fija en las capas de entrada y salida, definida así de forma a priori por que estamos tomando en cuenta las restricciones que plantea el problema particular de clasificación de patrones de estados de desarrollo celular. Así cada elemento de la población inicial de CNAs tiene cinco neuronas en la capa de entrada y tres neuronas en la capa de salida. El número de neuronas en la capa de entrada y de salida esta determinado por que en la población de patrones de estados celulares, cada patrón tiene cinco parámetros que distinguen las tres clases de datos a diferenciar, basura, mitosis y núcleos. En la figura 2.3 se muestra la configuración estructural que tienen los CNAs de nuestra propuesta metodológica.

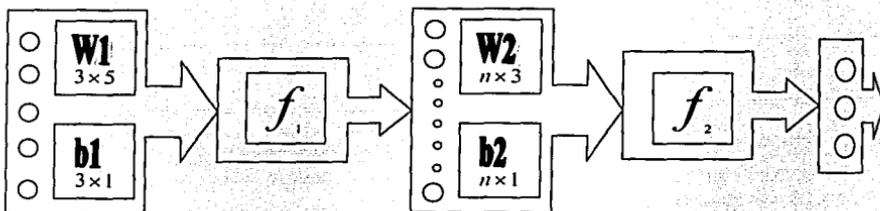


Figura 2.3 Estructura genotípica de los CNAs para el desarrollo y entrenamiento de clasificadores de patrones de estados celulares

El cromosoma genérico que estamos mostrando en la figura 2.3 nos esquematiza la forma genotípica que van a tener los CNAs en la población inicial. El número de neuronas en la capa intermedia es una variable de diseño y a la vez de entrenamiento. Nótese en la figura 2.3 que la dimensión de la capa intermedia no esta determinada. El rango de variación para este parámetro de diseño lo establecimos después de un proceso de ensayo y error, de modo que la capa intermedia puede tener entre 1 y 32 neuronas, cotas mínimas y máximas que definimos para la poda y crecimiento neuronal en la capa interna. En el proceso de encontrar un rango adecuado para este parámetro también nos percatamos que un número entre 2 y 25 neuronas en la capa intermedia es un intervalo útil de crecimiento y poda neuronal, pero decidimos dejar el intervalo inicial de 1 a 32 para explorar posibles configuraciones singulares de CNAs.

### 2.3.2 MATRIZ DE CONFUSIÓN Y FUNCIÓN MULTIOBJETIVO

¿Cómo será evaluado el desempeño de clasificación de los CNAs iniciales para producir nuevos CNAs mejorados, con buenos porcentajes de clasificación y sin tendencias o sesgos de clasificación? La mayoría de los algoritmos clásicos de entrenamiento neuronal hacen uso del error cuadrático [HAY98], y sus derivados el error cuadrático medio y promedio, como medida del desempeño de la red neuronal en base al ajuste de los pesos. Nuestro problema es encontrar buenos CNAs, en específico CNAs no sesgados a una clase determinada y el error cuadrático promedio es una expresión puramente matemática que no nos permite percibir información del desarrollo estadístico de las clasificaciones que está realizando la red neuronal durante la etapa de entrenamiento. Así entonces debemos emplear un tipo de medida no puramente cuantitativa, lo que necesitamos es una expresión que nos diga más acerca de que tan bien o mal esta clasificando un CNA durante su ajuste de pesos, pero de forma cualitativa y cuantitativa. Lo que estamos diciendo es proponer una expresión que combine criterios lógicos y claro estadísticos, pero en el sentido estadístico explícito que exige una clasificación. Estamos induciendo una expresión multicriterio lógico-matemática. Para este fin vamos a utilizar una matriz de confusión la cuál se muestra en la tabla 2.7 y se describe a continuación [MAS93].

Distribuciones		Resultados				
Clases	# de patrones	Clase1	Clase2	Clase3	# de Errores	Frecuencia relativa de Aciertos
Clase1	115	102	0	13	13	<b>0.8870</b>
Clase2	232	1	211	20	21	<b>0.9095</b>
Clase3	271	27	23	221	50	<b>0.8155</b>
Total	618	130	234	254	84	<b>0.8641</b>
Frecuencia relativa de Falsos positivos		<i>0.2154</i>	<i>0.0983</i>	<i>0.1299</i>		
Suma del error cuadrático: 34.130000						
Error cuadrático promedio: 1.268587						

Tabla 2.7 Matriz de confusión resultado de la clasificación de un CNA que clasifica tres clases

Supongamos que la tabla 2.7 muestra los resultados de clasificación de una red neuronal artificial que clasifica un conjunto de datos separados en tres clases, en la primer columna podemos ver el número de datos que tiene cada clase dentro de la muestra, los elementos  $a_{1,1}$ ,  $a_{2,1}$  y  $a_{3,1}$ , clase1, clase2 y clase3 cuyos valores son 115, 232 y 271, respectivamente. El

cuarto renglón de la primer columna, el elemento  $a_{1,4}$  con el valor 618, es el total de los datos de la muestra que se está clasificando. En la última columna de la matriz podemos ver la frecuencia relativa de clasificaciones que alcanzó la red neuronal para cada clase, elementos  $a_{1,6}$ ,  $a_{2,6}$ , y  $a_{3,6}$ , para este caso, con valores 0.8870, 0.9095, y 0.8155, respectivamente, y el elemento  $a_{4,6}$  es la frecuencia relativa total de aciertos de clasificación de la red neuronal, cuyo valor es 0.8641. En el último renglón de la tabla 2.7 podemos ver tres elementos. En la matriz de confusión son los elementos  $a_{5,2}$ ,  $a_{5,3}$ ,  $a_{5,4}$ , con valores 0.2154, 0.0983, 0.1299, respectivamente. Estos elementos son las frecuencias relativas de clasificaciones erróneas tomando en cuenta el número de clasificaciones correctas; es decir, son los valores de falsos positivos de clasificación. Cuanta información estadística y sobre todo funcional hemos podido extraer de la matriz de confusión del CNA, lo que no hubiéramos podido hacer tomando sólo en cuenta el error cuadrático promedio, que en el caso de la matriz de confusión de la tabla 2.7 es igual a 1.268587.

Utilizando la matriz de confusión, en específico los elementos que hemos mencionado, vamos a construir nuestra función multiobjetivo. Debemos tener en mente que uno de nuestros objetivos es que los CNAs de nuestra población deben mantenerse lo menos sesgados posible. Otro objetivo es el compromiso de que se mantengan en una región factible de clasificación. Esto es, debemos exigir que los CNAs no dominados clasifiquen dentro de ciertos intervalos de frecuencias relativas de clasificaciones correctas. Los intervalos o región factible que estamos proponiendo son  $[.8,.99]$ , es decir los buenos CNAs en la población son los que clasifiquen con un porcentaje de certeza superior al 80% e inferior al 99%, con objeto de que nuestros CNAs no sean bajo ni sobre entrenados. Mas no solo deseamos que nuestros CNAs clasifiquen bien sino que además queremos que el porcentaje de falsos positivos sea el menor posible, así que nuestra función objetivo debe exigir también que en cada generación del algoritmo genético el porcentaje de falsos positivos se vaya minimizando. Por último, para garantizar que en las poblaciones de clasificadores los CNAs sesgados a una clase determinada no preserven su herencia, vamos a exigir que la función multiobjetivo tome en cuenta la dispersión de las clasificaciones correctas mediante la medida de la varianza de los elementos  $a_{1,6}$ ,  $a_{2,6}$  y  $a_{3,6}$ . Así, aseguramos que nuestros clasificadores no van a clasificar sólo bien una o dos de tres

clases, sino que un CNA no dominado o bien adaptado de una población, clasificará bien las tres clases conjuntamente.

La función multiobjetivo a minimizar que estamos proponiendo, utilizando los elementos de la matriz de confusión que hemos mencionado es la siguiente:

$$F_{obj} = ((.8 < a_{1,6} < .99, .8 < a_{2,6} < .99, .8 < a_{3,6} < .99), (a_{5,2} + a_{5,3} + a_{5,4}), \text{var}(a_{1,6}, a_{2,6}, a_{3,6}))$$

Ecu 2.2

En resumen la ecuación 2.2 plantea de manera simultánea:

- En los primeros tres términos el objetivo es maximizar los valores de las tres frecuencias relativas de aciertos de clasificación, restringidos por tres funciones de membresía determinadas respectivamente en tres intervalos de clasificaciones correctas
- En el cuarto término el objetivo es minimizar una función puesta en términos de la sumatoria de las frecuencias relativas de falsos positivos. Evitar en lo general cometer errores de clasificación.
- En el quinto término el objetivo es minimizar una función puesta en términos de la varianza cuyos argumentos son las frecuencias relativas de aciertos de clasificación. Evitar el sesgo en el CNA.

Así la ecuación 2.2 tiene más información de las frecuencias relativas o porcentajes de clasificación, de las restricciones de clasificación y de los objetivos de clasificación para la selección de buenos CNAs que la expresión del error cuadrático promedio siguiente:

$$E = \frac{1}{n} \sum_{j=0}^{n-1} (t_{p,j} - o_{p,j})^2$$

Ecu 2.3

La cual sólo toma en cuenta el promedio de la diferencia entre los valores deseados  $t_{p,j}$  (targets) y las respuestas  $o_{p,j}$  del CNA.



A este respecto existen algunas otras expresiones o normas alternativas que pueden sustituir al Error Cuadrático Promedio para evaluar el desempeño de una Red Neuronal [KUR00b], [KUR01], [YAO97]. Pero, considerando que un CNA es un caso particular de red neuronal, es un objeto que debe ser evaluado con métricas estadísticas más complejas, como las que muestra la ecuación 2.2.

### 2.3.3 MULTIPOBLACIÓN Y MIGRACIÓN

Como se dijo en el capítulo uno, cuando se esta hablando de una optimización multiobjetivo las simulaciones han demostrado que el algoritmo genético simple no es capaz de mantener, de forma óptima, el conjunto de soluciones factibles en más de un objetivo simultáneamente. A este fenómeno se le llama deriva genética o amontonamiento genético, ya que significa que la población tiende a converger y a concentrarse alrededor de una u otra alternativa de optimización. Esto se debe a dos factores:

- No se asegura la dispersión poblacional necesaria para alcanzar todos los objetivos simultáneamente.
- Existe un error aleatorio asociado a los operadores genéticos [DEB89].

Este problema se acentúa cuando se utiliza una única población finita. En nuestro problema lo podemos notar como si fuera un fenómeno de endogamia o replicación genotípica. Cuando se utiliza una población única de CNAs, los mejores elementos tienden a dominar y a replicarse en exceso en toda la población. Esto provoca que el desarrollo de la población se corrompa en las generaciones medias del algoritmo genético resultando en CNAs sesgados; es decir, que clasifican muy bien una determinada clase, pero el resto de las clases son reconocidas con porcentajes bajos de aciertos. Entonces debido a que fue necesario que planteáramos evaluar los CNAs mediante una función multiobjetivo, es importante evitar que la población de CNAs tenga una tendencia hacia la concentración a uno u otro objetivo de optimización. Esto es muy importante, debido a que mantener valores óptimos de forma simultánea en todos los objetivos, significará una propiedad importante que deben cumplir los CNAs que entrene y diseñe nuestra metodología; es

decir, queremos encontrar CNAs con buenos porcentajes de clasificación y que no tengan sesgo o tendencia de mejores porcentajes de clasificación hacia una clase determinada. Así se vuelve una restricción, que los porcentajes de clasificación de los CNAs tengan una varianza muy baja entre ellos, y a la vez que todos los porcentajes de clasificación estén dentro de un intervalo definido como intervalo aceptable de clasificación. Como dijimos en el capítulo uno, la técnica de formación de nichos basada en los esquemas de densidad o repartición (crowding and sharing scheme)[DEB89] no funcionan en nuestro caso particular de prueba, ya que para ambos esquemas no tenemos de forma analítica la función a optimizar y por lo tanto no contamos con la forma de la superficie de la función multicriterio a optimizar, ni mucho menos tenemos información del número de picos (mínimos o máximos globales) con los que cuenta tal superficie de desempeño. En el caso del desarrollo y entrenamiento de CNAs, la superficie de la función objetivo se va construyendo con las evaluaciones de cada CNA en la población.

Nuestra propuesta se orienta en particionar la población de CNAs en tres familias independientes de CNAs, dejarlas evolucionar independientemente una de la otra. Durante algunas generaciones del Algoritmo Genético, seleccionadas de forma aleatoria, vamos a realizar perturbaciones aleatorias en el desarrollo individual de las familias de CNAs mediante un proceso de migración sin restricciones entre las tres familias, con objeto de provocar variabilidad en la evolución de los CNAs y así evitar la deriva genética. Así estamos provocando que exista una cierta medida de dispersión necesaria para poder alcanzar todos los objetivos de clasificación de forma simultánea. De forma esquemática se resumen estas ideas en la figura 2.4. En esa figura podemos ver que no nos alejamos mucho de la estructura general del algoritmo genético simple. En nuestra propuesta particular inicialmente generamos tres familias de CNAs, cada familia tiene 50 individuos. Este número de individuos por familia lo obtuvimos de manera experimental y nos dimos cuenta que después de este número de CNAs, los resultados de clasificación dejan de ser sensibles a este parámetro. Otro criterio para seleccionar 50 CNAs por familia fue que cada CNA es un elemento que consume tiempo de procesamiento de máquina lo que redundaba en el desempeño del algoritmo en general.

TESIS CON  
FALLA DE ORIGEN

# TESIS CON FALLA DE ORIGEN

## CAPÍTULO DOS

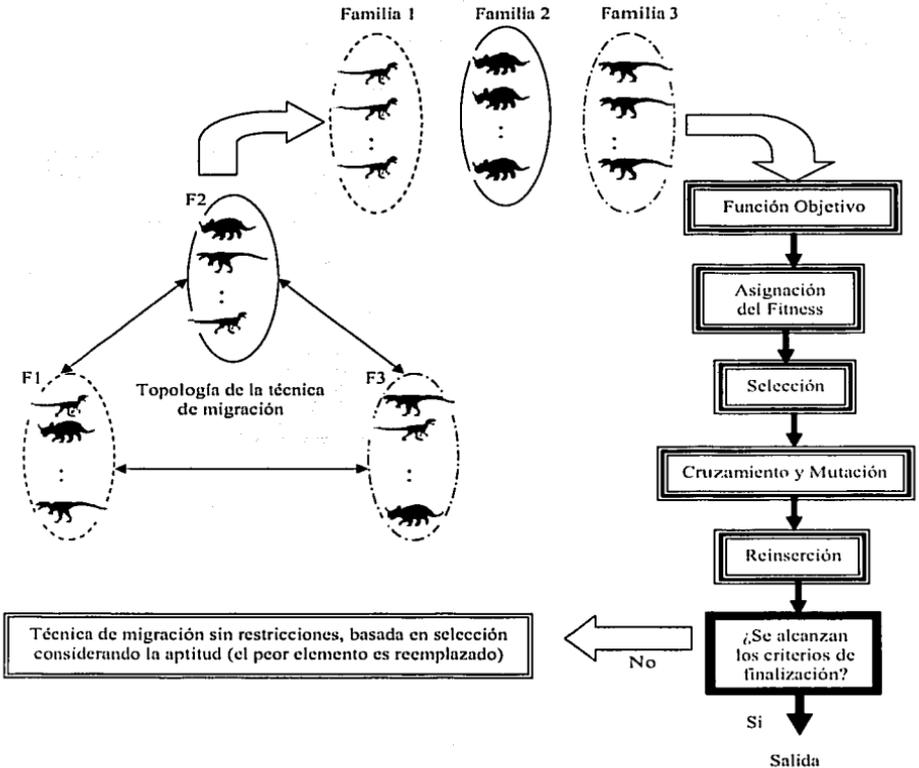


Figura 2.4 Esquema del Algoritmo Genético Multipoblacional con mecanismo de migración sin restricciones

La variante al algoritmo genético simple que estamos proponiendo se muestra en la figura 2.4. Este hace uso de tres familias de CNAs generadas inicialmente de forma aleatoria y que se desarrollan independientemente. Esto con el fin de evitar el fenómeno de endogamia o replicación genotípica, ya que cuando se utiliza una única población de CNAs, los

mejores adaptados tienden dominar y a replicarse en exceso en toda la población, provocando que el desarrollo de la población total se corrompa resultando en clasificadores sesgados, con sobre o bajo entrenamiento.

Con el uso de familias separadas e independientes, lo que logramos es que en alguna generación intermedia del algoritmo genético, cada familia de CNAs se desarrolle y alcance cierta madurez de clasificación, independiente de las otras familias de CNAs. Posteriormente en algunas generaciones aleatoriamente determinadas, se lleva a cabo un mecanismo similar a la migración no restringida entre poblaciones para provocar una perturbación en la evolución de cada población y asegurar de esta manera la dispersión poblacional y consecuentemente en cierta medida asegurar la dispersión a cumplir los objetivos de clasificación de forma simultánea. Esta idea tiene su analogía en las teorías de nichos y la formación de especies en los modelos biológicos. Cada familia de CNAs se comporta como una especie distinta que se está desarrollando en un nicho determinado, en donde se está llevando a cabo de manera aislada la selección natural, la lucha por la sobrevivencia de la herencia de los individuos mejor adaptados, esto es, individuos dominantes en una familia tratan a toda costa de mantener su dominio que les asegura la continuidad de su descendencia. Tal continuidad de la herencia solo se ve amenazada cuando se presenta la migración de nuevos individuos, pertenecientes a otras familias que compiten con los individuos de la familia receptora (exclusión natural entre individuos dominantes) y favorecen la diversidad de las características hereditarias y por lo tanto, la variación y mejoramiento de la especie y de la población en general.

Es importante decir que, dado que estamos utilizando un mecanismo de migración sin restricciones, esto no significa que todos los individuos de una familia pueden migrar a cualquier familia, es necesario cuidar las veces y el número de migrantes. Elegimos un porcentaje de migración del 50% en base al valor de aptitud. Esto es que 50% de los CNAs de una familia se eligen de forma aleatoria y se reemplazan por una selección de los CNAs mejor adaptados de las otras familias. Este mecanismo de migración se realiza con una probabilidad del 30% en cada generación del plan evolutivo. Esta combinación de un porcentaje relativamente alto de individuos a migrar con un porcentaje relativamente bajo de probabilidad de que ocurra la migración se consideró tomando en cuenta que si bien utilizamos el mecanismo de migración como un mecanismo perturbador del desarrollo

individual de cada familia de CNAs, no queremos excedernos y evitar que en lo general el mecanismo de migración se convierta en una especie de artificio que revuelve o mezcla a las familias en el algoritmo genético y terminemos teniendo a fin de cuentas una sola población partida en tres conjuntos que está siendo “barrajada” de forma irregular.

### 2.3.4 ASIGNACIÓN DEL VALOR DE APTITUD RELATIVO

Ya definimos la estructura cromosómica no homogénea más adecuada que va a caracterizar a los CNA, también establecimos cual es la manera más conveniente de evaluar el desempeño de los CNAs y la definimos de manera que ocupa más de una dimensión. Pero la función multiobjetivo al ser puesta en términos vectoriales no nos permite establecer una relación de orden. Establecer una evaluación y posteriormente una jerarquía del desempeño, tiene una importancia que se remarca en las técnicas evolutivas, por el hecho de que se trabajan con una población o un conjunto de subpoblaciones finitas, cuyos elementos significan aproximaciones o soluciones potenciales. A éstas se les aplica iterativamente el principio de selección de las mejores soluciones obtenidas mediante la evaluación de la función objetivo determinada por la naturaleza de nuestro problema a resolver; apostando que tal principio de selección aproxime cada vez más a la solución óptima. Así, en cada iteración, un nuevo conjunto de aproximaciones es generado y se aplica de nuevo el proceso de selección de las mejores soluciones. Por este procedimiento resulta importante saber cuales son los mejores CNAs dentro cada familia, entonces necesitamos asignar a cada CNA un valor de aptitud relativo a la familia que pertenece el CNA. Para asignar el valor de aptitud, elegimos la técnica de Fonseca y Flemming[FON95]. Ellos proponen una variación a la técnica clásica de Pareto[GOL89]. En el esquema de Fonseca la aptitud de un individuo corresponde con el número de individuos en la población actual por los cuales es dominado. Los individuos no dominados tienen la misma aptitud, mientras que los dominados son castigados o penalizados de acuerdo a que tan concentrada está la población en el correspondiente conjunto de elementos dominados. En la figura 2.5, podemos observar como se asigna la aptitud con la técnica de Fonseca y Flemming, en la cual la población es ordenada de acuerdo al grado de dominación expresado como:

$$rank(x_u, t) = r_u^{(t)}$$

Ecu. 2.4

En donde el individuo  $x_{ii}$ , con su correspondiente vector objetivo  $u_i$ , en la generación  $i$  es dominado por  $r'_i$  individuos en la población actual.

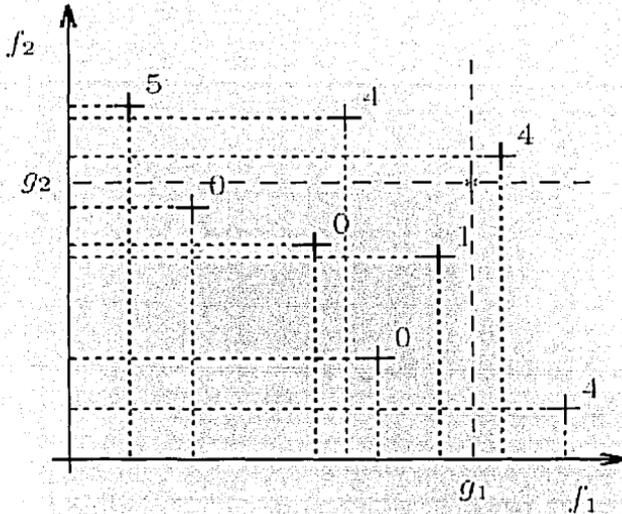


Figura 2.5 Asignación de la jerarquía de desempeño o aptitud relativa de acuerdo con la técnica de Fonseca y Fleming

Compárese la figura 2.5 con la figura 1.3 del capítulo uno, donde se muestra la asignación de la aptitud relativa mediante la técnica de Pareto. La dominancia de Fonseca nos permite distinguir de manera más refinada a los elementos dominantes que la técnica clásica de Pareto, en la cual ciertos CNAs, específicamente los CNAs que se encuentran en los extremos laterales del frente de Pareto, pueden pasar por dominantes sin serlo de forma estricta [DEB01], [ROS03]. Para nuestra aplicación, un CNA que se sitúe en los extremos laterales del frente de Pareto significa un CNA sesgado. Además como podemos notar de la figura 2.5, la técnica de Fonseca permite en mayor medida que elementos con un

desempeño promedio en todas las dimensiones, tengan una jerarquía más acorde con su desempeño y no sean confundidos con elementos con mal desempeño en general.

La dominancia de Fonseca, al ser una técnica basada en Pareto, se debe usar en conjunción con las técnicas de formación de nichos y creación de especies para mantener apropiada la diversidad en la población[DEB89]. Las técnicas estándar de nichos y especies requieren que se tenga una estimación del número de picos que tiene la función objetivo. Este hecho en nuestro caso es imposible, dado que estamos tratando con una función objetivo que no está descrita completamente; el diseño y entrenamiento de CNAs no permite el uso de una función analítica que permita estimar el número de picos en ella. Para reducir los efectos de la deriva genética nuestra propuesta orienta un cambio fundamentado en la combinación de una técnica basada en Pareto, la dominancia de Fonseca, con el mecanismo de migración y manejo de multipoblaciones, descritos anteriormente.

### 2.3.5 ASIGNACIÓN DE LA PROBABILIDAD DE SELECCIÓN PROPORCIONAL AL VALOR DE APTITUD RELATIVO

Basándonos en la matriz de confusión, determinamos de forma vectorial la manera más adecuada para evaluar el desempeño de los CNAs. Posteriormente con el vector que caracteriza el desempeño, mediante la dominancia de Fonseca asignamos el valor de aptitud a cada CNA relativo a la familia que pertenece. Estas medidas para calificar absoluta y relativamente a los CNAs son necesarias para que finalmente se pueda asignar a cada CNA un valor de probabilidad de que sus características genotípicas y fenotípicas se preserven en las generaciones sucesivas del plan evolutivo. Existen varias técnicas para asignar la probabilidad de sobrevivencia, la mayoría manejan el concepto de ranqueo (*rank*) el cual se refiere a una posición determinada por el valor de aptitud relativo (la población se ordena de tal forma que el individuo con menor valor de aptitud relativo tiene la posición o ranqueo cero y el individuo con el mayor valor de aptitud relativo se le asigna el ranqueo o la posición  $(n - 1)$ , para una población de tamaño  $n$ ). Realizamos pruebas con dos tipos muy conocidos de ranqueo: el ranqueo lineal y el no lineal[BAC97]. En el Ranqueo Lineal la probabilidad de selección para el individuo  $i$  es:

$$P_{lin\_rank}(i) = \frac{\alpha_{rank} + \left[ \frac{rank(i)}{(n-1)} \right] (\beta_{rank} - \alpha_{rank})}{n} \quad \text{Ecu. 2.5}$$

donde  $\alpha_{rank}$  y  $\beta_{rank}$  son el número de descendientes asignados en cada generación para el peor y mejor individuo, respectivamente.

En el Ranqueo no Lineal, de forma similar que en el lineal, se asigna la probabilidad de selección en base a la jerarquía (ranking) o dominancia en nuestro caso. Pero esta técnica, al ser no lineal, no es proporcional al ranqueo, y está determinada por la siguiente ecuación:

$$P_{nolin\_rank}(i) = \frac{\alpha + \left[ \frac{rank(i)^2}{(n-1)^2} \right] (\beta - \alpha)}{c} \quad \text{Ecu. 2.6}$$

donde  $c = \frac{(\beta - \alpha)n(2n-1)}{6(n-1)} + \alpha n$  es un factor de normalización y  $\alpha$  y  $\beta$  determinan el

intervalo de selección de la forma  $\left[ \frac{\alpha}{c}, \frac{\beta}{c} \right]$ , de tal manera que  $0 < \alpha < \beta$ .

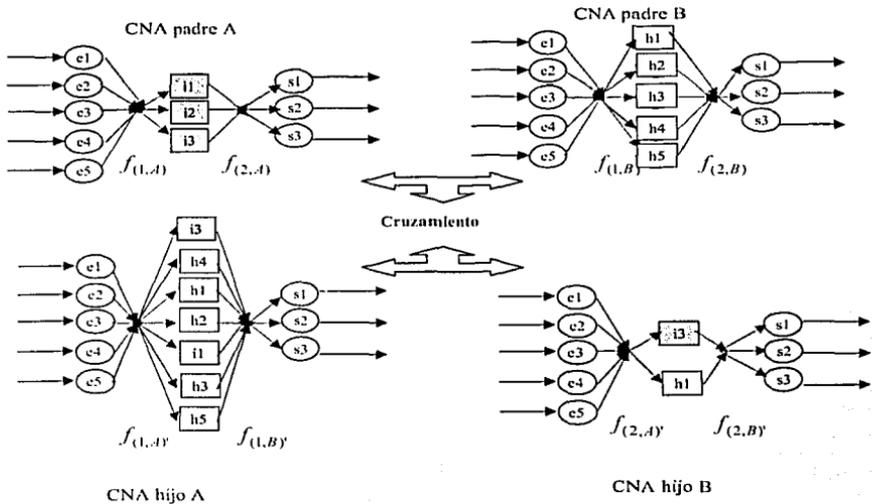
Observamos de forma experimental que el ranqueo lineal a comparación con el ranqueo no lineal, tiene mayor tendencia a una prematura convergencia y por consiguiente a caer en mínimos locales. Así que la asignación de la probabilidad de selección proporcional al valor de aptitud relativo, o más brevemente llamado ranqueo, lo establecimos mediante la expresión no lineal mostrada en la ecuación 2.6, cuyos valores  $\alpha$  y  $\beta$  los definimos en 0.1 y 2, respectivamente.

Para seleccionar cuales CNAs van a heredar sus características a la siguiente generación, en nuestro enfoque utilizamos el muestreo estocástico universal[CHI95], el cual selecciona un número determinado de CNAs en cada familia, para reproducirse de acuerdo a su valor de probabilidad de selección proporcional al valor de aptitud relativo a la familia o ranqueo. En nuestro enfoque establecimos un porcentaje de brecha generacional igual al 90%, es decir, en cada familia 90% de los CNAs van a ser seleccionados, según sus probabilidades de selección, para mutar y reproducirse. El porcentaje de selección antes mencionado lo

elegimos porque queremos que la mayoría de cada familia de CNAs participe en el desarrollo de los mejores CNAs pero que tal participación también la determine el valor de aptitud relativo de cada CNA. Este proceso de selección es completamente probabilístico lo cual nos asegura en alguna medida que, no existirá alguna otra tendencia de selección para determinados CNAs que no sea el desempeño individual de cada CNA dentro de cada familia. Además dentro de la probabilidad que maneja el muestreo estocástico universal, cabe la posibilidad que se seleccionen algunos CNAs cuyos porcentajes de clasificación sean regulares incluso, malos y esto conviene al plan evolutivo, ya que no es garantía que de la selección y posteriormente cruzamiento y mutación de solo elementos buenos ( no dominados ) resulten en la obtención de CNAs no dominados. Incluso se puede corromper la descendencia con una selección muy elitista, sin una dosis sana de aleatoriedad. En la variedad está gran parte de la solución. La selección, cruzamiento y mutación de CNAs dominados y no dominados trae consigo variedad en la descendencia y alta probabilidad de encontrar CNAs no dominados con buenos porcentajes de clasificación. Es por esto que el muestreo estocástico universal es conveniente usarlo en nuestra propuesta metodológica.

### 2.3.6 OPERADORES DE CRUZAMIENTO Y MUTACIÓN

Hasta aquí ya tenemos determinados porqué y que CNAs van a preservar sus características fenotípicas y genotípicas en las sucesivas evoluciones de la metodología que estamos proponiendo. Pero ¿de qué manera se llevarán acabo los mecanismos para el intercambio de la información de tales características? los mecanismos de intercambio de información están determinados mediante los operadores de cruzamiento y mutación. Como nosotros determinamos nuestra propia estructura cromosómica, los operadores de cruzamiento y mutación deben estar diseñados para permitir el intercambio de características de diseño estructural y valores de magnitud en las matrices de pesos, vectores de tendencia, además de permitir el intercambio de forma aleatoria de la configuración de las funciones de activación de los CNAs. Como se enunció en la introducción, las propuestas de implementación del operador de cruzamiento fueron varias. De la conjugación de tales propuestas se realizaron algunas pruebas, produciendo finalmente una propuesta propia, el Cruzamiento Intermedio Neuronal Aleatorio de Valor y Estructura (CINAVE), el cual mostramos en la figura 2.6.



$$f_{(i,X)}, \in [f_{(1,A)}, f_{(1,B)}, f_{(2,A)}, f_{(2,B)}] \quad \forall i = 1, 2 \text{ y } X = A, B$$

Figura 2.6. Funcionamiento del Cruzamiento Intermedio Neuronal Aleatorio de Valor y Estructural (CINAVE) a nivel de cambio de la configuración estructural de los CNAs

En la figura 2.6 podemos ver como funciona nuestra propuesta de operador de cruzamiento. De forma aleatoria se seleccionan dos CNAs padres, se verifica que la suma de las dimensiones de las capas intermedias de los padres no exceda la cota máxima de neuronas en un CNA, la cual definimos como 32. Si no se excede tal cota entonces, de forma uniformemente aleatoria, se eligen dos números naturales del intervalo  $[1, \dim(CNA_i^A) + \dim(CNA_i^B)]$  donde  $\dim(CNA_i^A)$  significa la dimensión de la capa  $i$  del CNA A, en nuestro caso particular  $i = 2$  ya que los CNAs que manejamos tienen tres capas y la capa número 2 es la intermedia. Si la expresión  $\dim(CNA_i^A) + \dim(CNA_i^B) > 32$  con

$i = 2$  es cierta, entonces el par de números naturales aleatorios se elige del intervalo  $[1,32]$  con objeto que ningún CNA exceda la cota máxima de neuronas en la capa intermedia. La pareja de números naturales aleatoriamente seleccionados son las dimensiones de la capa interna que van a tener los dos CNAs hijos descendientes. Determinadas las dimensiones de las capas intermedias de los CNAs descendientes, se seleccionan las neuronas de intercambio en ambos padres, también mediante una distribución uniformemente aleatoria, y se procede a construir los dos CNAs hijos. Las funciones de activación de cada CNA hijo están determinadas por las funciones de activación que poseen los padres y se asignan de forma aleatoria, como se muestra en la figura 2.6.

La manera de construir las matrices de pesos y vectores de tendencia de los CNAs hijos, se muestra en la figura 2.7.

Matriz de pesos del CNA p

$$\begin{pmatrix} W^p_{1,1} & \cdots & W^p_{1,m_p} \\ \vdots & \vdots & \vdots \\ W^p_{n_p,1} & \cdots & W^p_{n_p,m_p} \end{pmatrix}$$

Matriz de pesos del CNA q

$$\begin{pmatrix} W^q_{1,1} & \cdots & W^q_{1,m_q} \\ \vdots & \vdots & \vdots \\ W^q_{n_q,1} & \cdots & W^q_{n_q,m_q} \end{pmatrix}$$

Matriz de pesos descendiente

$$\begin{pmatrix} h_{1,1} & \cdots & h_{1,m_s} \\ \vdots & \cdots h_{i_r,j_s} \cdots & \vdots \\ h_{n_r,1} & \cdots & h_{n_r,m_s} \end{pmatrix}$$

Donde  $h_{i_k,j_l}$  es un número uniformemente aleatorio escogido dentro del intervalo  $[\min(w^p_{i_s,j_s}, w^q_{i_s,j_s}), \max(w^p_{i_s,j_s}, w^q_{i_s,j_s})]$  y  $n_r \in \max(n_p, n_q), m_s \in \max(m_p, m_q)$

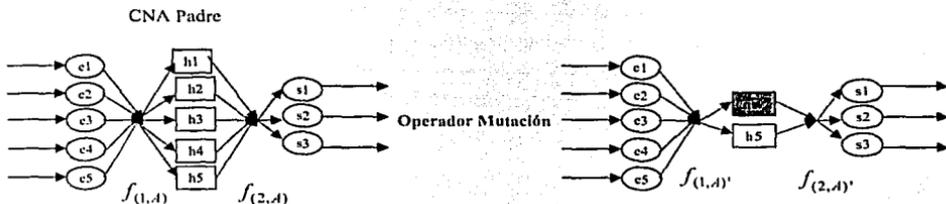
Figura 2.7. Funcionamiento del Cruzamiento Neuronal Estructural Aleatorio Sencillo a nivel de cambio de las matrices de pesos de los CNA's

Como podemos ver el operador de cruzamiento que estamos proponiendo, no solo provoca cambios estructurales en la configuración de la arquitectura de los CNAs, sino también desarrolla cambios en las magnitudes de los valores de las matrices de pesos y vectores de términos de tendencia. De tal forma que los procesos de poda y crecimiento neuronal que estamos generando con el CINAVE, son más que simples procesos de eliminaciones e inserciones aleatorias de neuronas. El objetivo es que las características genéticas, tanto en forma y valor, de los CNAs padres se transmitan a los CNAs descendientes en las generaciones sucesivas del proceso evolutivo, y que los efectos de la herencia genética resulten en una amplia variedad genotípica y fenotípica. Esta variedad servirá también para amortiguar los efectos de la endogamia o deriva genética que se acentúa cuando el intercambio de información genética es reducido a procesos simples de intercambio de características genotípicas. Además, dada la compleja representación con la que definimos nuestra estructura cromosómica genérica, podemos permitirnos plantear operadores genéticos complejos, ya que en nuestro enfoque no contamos con un paso de codificación-decodificación. Así, evitamos realizar la transformación de genotipo a fenotipo y viceversa, es decir, no necesitamos mecanismos de vectorización-desvectorización ya que el operador de cruzamiento que estamos proponiendo es matricial. Así la evaluación de los CNAs es directa. Esto significa ahorro de tiempo de ejecución, el cual decidimos mejor invertirlo en el diseño de un operador de cruzamiento que involucrara tanto el manejo estructural como el manejo matricial para la modificación de los CNAs, como se ve en la figura 2.7.

El operador de CINAVE toma dos CNAs padres para producir dos CNAs descendientes o hijos que poseen algunas características de valor y forma heredadas de los padres. Este es nuestro operador principal de intercambio de información, el operador sexual como se menciona en la bibliografía. Mas en los modelos evolutivos reales y en los problemas propios de la ciencia e ingeniería, se ha visto que algunas veces son determinantes para el desarrollo de la población, alteraciones de los valores de los genes de un solo cromosoma padre. Esto trae consigo una medida de asegurar cierta diversidad en el desarrollo de la población en las sucesivas generaciones del plan evolutivo. En nuestro planteamiento, si tomamos en cuenta a la mutación como el intercambio de información entre un elemento de la población con un elemento completamente ajeno a la población, el operador de mutación lo podemos plantear como un caso especial del CINAVE, entre un CNA de la población y

un CNA generado completamente independiente de la población. En la figura 2.8 podemos ver un ejemplo de la Mutación Intermedia Neuronal Aleatoria de Valor y Estructura (MINAVE) que estamos proponiendo, el cual toma como base el CINAVE. El operador de mutación que diseñamos, al igual que el operador de cruzamiento, realiza cambios tanto en estructura de configuración como en valor, modifica la estructura genotípica, altera los valores de las matrices de pesos y vectores de tendencia y la combinación de las funciones de activación del CNA hijo tomando en cuenta las características genotípicas y fenotípicas del CNA padre.

Las tasas de cruzamiento y mutación son parámetros genéticos que se determinaron de forma experimental. Se realizaron pruebas con combinaciones de porcentajes de probabilidad de cruzamiento y mutación del 90% y 10%, del 80% y 20%, del 70% y 30% y del 80% y 30%, respectivamente. La combinación que mejores resultados dio fue la que tiene el porcentaje probabilidad de cruzamiento del 70% y de mutación del 30%; valores que si bien no son los estrictamente usuales, fueron suficientemente adecuados para contribuir un poco en la atenuación de la deriva genética.



donde  $nw$  es una neurona originaria de un CNA generado completamente independiente al desarrollo del CNA Padre y  $f_{(i,A)} \in F_{Activ}$  con  $i = 1, 2$

Figura 2.8. Funcionamiento de la Mutación Intermedia Neuronal Aleatoria de Valor y Estructural (MINAVE) a nivel de cambio en la configuración estructural de los CNA's

### 2.3.7 REINSERCIÓN

Hasta aquí, de cada familia, tenemos a los nuevos CNAs descendientes, producto del cruzamiento y mutación de los CNAs con mejor desempeño. Para actualizar el estado y

mantener el tamaño individual de las tres familias de CNAs que maneja nuestra modificación al algoritmo genético simple, los nuevos CNAs tienen que ser reinsertados en las familias actuales; esto es, cada familia actual de CNAs se actualiza reemplazando a los CNAs padres con los CNAs hijos. Nuestra propuesta de reinsertación se basa en el valor de desempeño, así los descendientes con mejor desempeño de cada familia reemplazan a los ancestros con peor desempeño. Este mecanismo de reinsertación además nos asegura que para que un CNA sobreviva sucesivas generaciones, este debe ser lo suficientemente apto, en nuestro caso lo suficientemente imparcial, sin sesgo de clasificaciones y con porcentajes de clasificación aceptables. Esto significa que los CNAs no dominados, los dominantes que se consiguieron en una generación determinada, no se pierdan en las generaciones siguientes, sino es que ya se han encontrado mejores CNAs con porcentajes de clasificación superiores y con menor sesgo de clasificación, lo cual asegura que se continúa escalando en la búsqueda de los CNAs óptimos en cada familia.

#### 2.4 RESUMEN

En la búsqueda de CNAs óptimos, bien diseñados, sin sesgos y con buenos porcentajes de clasificaciones, nos encontramos con el problema de la deriva genética. Para atenuar los efectos de la deriva genética empleamos la técnica de asignación del desempeño de Fonseca, pero en la práctica observamos que ésta por sí sola no funciona y la teoría también recomienda hacer combinación de técnicas basadas en Pareto con formación de nichos o de especies. Pero como explicamos en las secciones anteriores, el problema de diseñar y entrenar CNAs no permite plantear una función analítica en donde podamos estimar un número de picos para poder distribuir de forma uniforme los nichos o especies que sean necesarios crear. Como solución a esta limitación, planteamos el uso de tres familias o subpoblaciones de CNAs y establecemos un mecanismo de migración sin restricciones que se lleva a cabo de forma aleatoria como artificio de perturbación poblacional. Así, a manera de analogía, el esquema de manejar familias independientes de CNAs y mezclarlas aleatoriamente mediante migración sin restricciones, se puede tomar como una variante del método VEGA, el cual no está basado en Pareto [COE99]. Nuestra variante de VEGA, si bien no garantiza formación de nichos y especies, sí permite que individuos con desempeño medio en todas las dimensiones de la función objetivo sobrevivan y posteriormente se

conviertan, dada la heurística que planteamos en la construcción de nuestra función objetivo, en individuos con un desempeño óptimo en todas las dimensiones de la función objetivo.

Esta combinación de técnicas basadas y no basadas en Pareto constituyen nuestra principal aportación empírica a la optimización multiobjetivo basada en computación evolutiva, cuyo resultado se capitaliza en una metodología propia, la cual es resultado del análisis y la experiencia recopiladas en el desarrollo y entrenamiento de redes neuronales para el reconocimiento de patrones, la aproximación de funciones y problemas de clasificación que se nos han presentado. En particular, el problema de clasificar patrones estados de desarrollo celular obtenidos a partir de imágenes biomédicas es muy interesante, ya que tiene muchas características que lo hacen singular de tratar, como se notó en los capítulos uno y dos. En el capítulo tres, analizaremos los resultados obtenidos, y confirmaremos la eficacia de la metodología que proponemos, para el caso particular de clasificación de patrones estados de desarrollo celular.

## CAPÍTULO TRES

### EXPERIMENTACIÓN Y ANÁLISIS

#### 3.1 OBTENCIÓN DE RESULTADOS

En este capítulo mostramos los resultados obtenidos con la metodología que describimos en los capítulos anteriores. Nuestro enfoque es una metodología para el desarrollo y entrenamiento de Clasificadores Neuronales Artificiales (CNAs), basada en una técnica evolutiva, multiobjetivo y multipoblacional, la cual si hacemos a un lado que tiene una inspiración basada en una analogía simplificada de los conceptos de la teoría de la evolución natural, y nos centramos en que estamos tratando con un algoritmo que propone una búsqueda probabilística utilizando elementos aleatorios, estadísticos y probabilísticos. Entonces el problema se reduce a la búsqueda de una combinación de parámetros neuronales que está oculta en la naturaleza misma del problema de clasificación de patrones provenientes de imágenes biomédicas. La búsqueda de esa combinación de parámetros, tal como la planteamos nos establece de forma simultánea el diseño y ajuste de un conjunto de CNAs óptimos.

Como estamos hablando de una metodología fundamentada en una búsqueda probabilística, debemos tomar en cuenta varios factores importantes, entre los que destacan: el problema a resolver plantea un espacio de búsqueda cuya combinatoria es muy grande y compleja, el espacio de búsqueda es dependiente del conjunto a clasificar, y por lo tanto, no existe ninguna posibilidad de tener una aproximación de cualquier tipo de la superficie que describe la función objetivo. Además, el método de búsqueda es dependiente de algunos parámetros propios del método. Por tales razones tenemos que aceptar que no existe forma sencilla de asegurar la convergencia. Pero a cambio, la manera de poner a prueba el conjunto y conjunción de técnicas propuestas es mediante una amplia serie de experimentos con el fin de lograr obtener un análisis estadístico confiable para poder realizar afirmaciones que comprueben el funcionamiento de la metodología.

Como resultado del número de pruebas que realizamos, aproximadamente 300 ensayos, observamos que la fiabilidad de la metodología propuesta fue de aproximadamente 260 ensayos, los cuales dieron por resultado final conjuntos no dominados de CNAs, cuyas frecuencias relativas o porcentajes de clasificación son superiores al 80% de aciertos de

clasificación. Sabemos que 260 ensayos exitosos de 300 realizados no son prueba concluyente, en lo general, de que el método es fiable en un 86%, pero sí podemos afirmar que en este estudio, para el problema de prueba, los resultados muestran que el método da resultados satisfactorios con una probabilidad superior al 80%.

A continuación observaremos las gráficas de los porcentajes de aciertos de clasificación de los conjuntos finales de CNAs no dominados en las tres familias, en dos variantes: mediante el método de coordenadas paralelas, que describimos en un capítulo anterior y también mediante gráficas tridimensionales de los porcentajes aciertos de clasificación. El objeto de mostrar ambas gráficas es tener una panorámica visual detallada de los resultados obtenidos, además que del análisis gráfico de los resultados obtenidos, colateralmente se pueden apreciar relaciones entre los patrones de basura, mitosis y núcleos que no están definidas claramente en la población de patrones de etapas de desarrollo celular.

En el ámbito biológico y químico se afirma que el genotipo es el manual de instrucciones de un organismo, en el cual se encuentra la totalidad de la información genética; y que el fenotipo es el resultado visible de la interacción entre los genes, sus productos y el medio ambiente. De forma más abstracta y práctica, para nosotros el genotipo es la estructura genética explícita de una estructura cromosómica (estructura de datos), y el fenotipo es la expresión física del genotipo. Así junto con las gráficas de coordenadas paralelas y las gráficas tridimensionales, mostraremos las matrices de confusión ( fenotipos ) de los CNAs no dominados, útiles para evaluar su desempeño y para analizar las relaciones entre las tres clases de patrones celulares desde un punto de vista estadístico. Esta será una herramienta que nos ayudará a corroborar los resultados gráficos que se observen con las gráficas bidimensionales y tridimensionales.

En el análisis obviamente incluiremos la estructura final de algunos CNAs no dominados de las tres familias, con objeto de observar cuáles son las características de la configuración estructural ( genotipo ) de los CNAs no dominados finales y poder analizar la influencia de cada parámetro de diseño en el desempeño de los CNAs. Esto nos permitirá observar si el número de neuronas en la capa intermedia, si la selección de determinadas funciones de activación, o si cambios en las magnitudes de algunos elementos de las matrices de pesos o en los vectores de tendencia son factores determinantes para establecer una refinación del

diseño de CNAs para el caso específico de clasificación de patrones de etapas desarrollo celular.

En los resultados mostrados en este capítulo, en la mayoría de las figuras, se observará que mostramos algunos CNAs no dominados finales de cada una de las familias. Tal selección la realizamos sin tener en cuenta ningún criterio particular de discriminación, simplemente seleccionamos aleatoriamente algunos CNAs del conjunto que forman el frente de Pareto final de cada familia y estos son los que mostramos. Esto lo mencionamos a reserva de que tal selección quizá no elija a los mejores CNAs, o a los CNAs representativos del conjunto final de óptimos de Pareto de CNAs, pero como lo refiere la literatura y lo confirma la práctica, la decisión final debe hacerse fundamentándose en un juicio de valor entre las alternativas disponibles para alcanzar una decisión particular. Para el caso particular del diseño y entrenamiento de CNAs para el problema de clasificación de patrones de etapas de desarrollo celular, todos los CNAs no dominados encontrados son suficientemente buenos, lo cual le da estabilidad y solidez a la solución del problema general del diseño y entrenamiento de CNAs, ya que no se ofrece una solución única, como tradicionalmente se hace, sino que se proporciona un conjunto de soluciones.

En todas las pruebas particionamos el conjunto experimental de patrones de etapas de desarrollo celular en dos muestras: una de entrenamiento y una de validación. La muestra de entrenamiento siempre fue seleccionada con el método que se explicó en el capítulo dos, el cual asegura la representatividad de la muestra basándose en las contenciones en los intervalos típicos muestrales de cada parámetro de los patrones de etapas de desarrollo celular. En el capítulo dos, se mostró que la cardinalidad de la muestra de entrenamiento no es factor al que sea sensible la metodología de selección. Con esta facilidad, en cada prueba seleccionamos distintas muestras de entrenamiento, algunas con distintas cardinalidades y con distintas magnitudes de representación proporcional de las tres clases de patrones de etapas de desarrollo celular, mas todas ellas representativas del conjunto experimental de patrones de etapas de desarrollo celular.

### 3.2 ANÁLISIS

Como dijimos en el capítulo dos, elegimos como técnica de muestreo particionar la población experimental de patrones de estados de desarrollo celular en dos muestras: una

de entrenamiento y una de validación. Los nombres de ambas muestras son perfectamente indicativos de su función. La muestra de entrenamiento siempre va a poseer tres características:

- La cardinalidad de la muestra de entrenamiento es menor que la cardinalidad de la muestra de validación.
- La muestra de entrenamiento no contiene ningún elemento de la muestra de validación, ya que junto con la muestra de validación, componen una partición de la población experimental de patrones de etapas de desarrollo celular.
- Y sobre todo la muestra de entrenamiento es representativa de la población experimental, claro, la representatividad esta asegurada mediante el método de selección alatorio tomando en cuenta los intervalos típicos descritos en el capítulo dos.

Entonces, el complemento de la muestra de entrenamiento es la muestra de validación, y como ya se dijo tiene una cardinalidad mayor. Además de esta característica ninguna otra consideración especial se le exige, ya que a la muestra de validación no se le demanda ser representativa, no importa, esa es característica esencial para la muestra de entrenamiento. El papel importante de la muestra de validación es probar realmente la capacidad de los CNAs obtenidos con nuestra propuesta de metodología de desarrollado y diseño evolutivo y multiobjetivo. La prueba importante de todo clasificador neuronal se realiza cuando se le presentan a clasificar nuevos patrones, los cuales nunca se le presentaron en la etapa de ajuste o entrenamiento. Es en la etapa de clasificación con la muestra de validación, propiamente llamada etapa de validación, en donde se realiza la verdadera prueba del desempeño de los CNAs no dominados hallados, porque en todo el desarrollo y entrenamiento evolutivo de las tres familias de CNAs, en cada generación, se evaluaron fenotípicamente todos los CNAs con la muestra de entrenamiento. De tal etapa de entrenamiento, es de esperarse que con la muestra de validación los CNAs obtenidos den buenos porcentajes de clasificaciones correctas, ya que a la muestra de entrenamiento se le exige ser representativa. Además, de que la metodología trata de construir y ajustar los CNAs de forma óptima, evitando el sesgo que provoca el bajo y sobre entrenamiento y que

se expresa en bajos porcentajes de clasificaciones correctas o en frecuencias relativas de clasificaciones que en conjunto tienen una varianza alta. Esa problemática al igual que en los métodos de entrenamiento neuronal clásicos se presenta también en los métodos evolutivos multiobjetivo pero más asociada con la deriva genética que con la calibración del número de neuronas, del número de capas en el CNA y de la elección del número de iteraciones de entrenamiento o epochs[SM193].

### 3.2.1 DESARROLLO Y EVOLUCIÓN CON UNA MUESTRA HOMOGÉNEA

En esta sección vamos mostrar algunos resultados de nuestra metodología, cuando se seleccionaron muestras de entrenamiento homogéneas en las distribuciones de las proporciones de cada clase de patrones y representativas de la población de patrones de etapas de desarrollo celular.

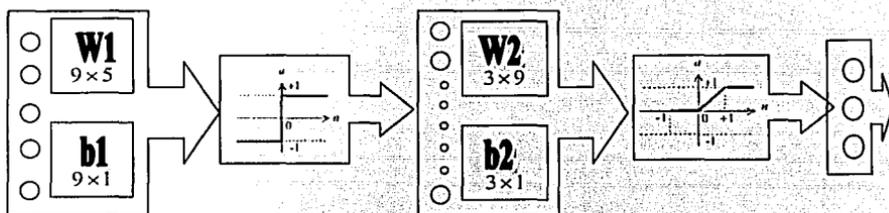


Figura 3.1 Configuración de la arquitectura de un CNA no dominado de la familia uno de CNAs

En la figura 3.1 podemos observar la configuración de la arquitectura de un CNA, tomado al azar, perteneciente al conjunto de CNAs no dominados de la familia uno. Genotípicamente el CNA no dominado está bien construido: tiene función de activación escalón simétrica en la capa intermedia, nueve neuronas en la capa intermedia, y en la capa de salida función lineal saturada. Genotípicamente este CNA no dominado no posee ninguna singularidad en su configuración, quizá sea relevante subrayar que la combinación de tales funciones de activación está de acuerdo con los valores de las salidas deseadas de los patrones ya que la función escalón simétrica en la capa intermedia mapea a -1 y 1, todos los valores de las respuestas de las neuronas intermedias. Posteriormente, dependiendo de:

la combinación de valores de los pesos, términos de tendencia asociados a la capa de salida, y de las respuestas de las neuronas de la capa intermedia; la función lineal saturada de la capa de salida, mapea todos los valores de las respuestas de las neuronas de salida a valores en el intervalo  $[0,1]$ . Esto está de acuerdo con los valores de los vectores de las salidas deseadas de los patrones de etapas de desarrollo celular de la muestra de entrenamiento. Este análisis demuestra que genotípicamente el CNA no dominado seleccionado corresponde con los objetivos establecidos inicialmente y está estructuralmente bien diseñado. Extendamos este análisis hacia el enfoque funcional, es decir, al punto de vista de su desempeño.

Distribución de los patrones		Resultados del Clasificador				
Clases	# de patrones	Mitosis	Núcleos	Basura	# de Errores	Frecuencia relativa de Aciertos
Mitosis	55	52	0	3	3	<b>0.9455</b>
Núcleos	55	0	54	1	1	<b>0.9818</b>
Basura	55	3	1	51	4	<b>0.9273</b>
Total	165	55	55	55	8	<b>0.9515</b>
Frecuencia relativa de Falsos positivos		<b>0.0545</b>	<b>0.0182</b>	<b>0.0727</b>		
Suma del error cuadrático: 34.130000						
Error cuadrático promedio: 0.206848						

Tabla 3.2a Matriz de confusión resultado de la clasificación con la muestra de entrenamiento

Distribución de los patrones		Resultados del Clasificador				
Clases	# de patrones	Mitosis	Núcleos	Basura	# de Errores	Frecuencia relativa de Aciertos
Mitosis	127	114	0	13	13	<b>0.8976</b>
Núcleos	276	4	242	30	34	<b>0.8768</b>
Basura	332	42	11	279	53	<b>0.8404</b>
Total	735	160	253	322	100	<b>0.8639</b>
Frecuencia relativa de Falsos positivos		<b>0.2875</b>	<b>0.0435</b>	<b>0.1335</b>		
Suma del error cuadrático: 165.069000						
Error cuadrático promedio: 0.224584						

Tabla 3.2b Matriz de confusión resultado de la clasificación con la muestra de validación

Figura 3.2 Resultados de clasificación del CNA no dominado de la figura 3.1

En la figura 3.2 podemos ver dos tablas, corresponden a las matrices de confusión resultado de la operación de clasificación del CNA no dominado de la figura 3.1. En la tabla 3.2a se

muestra la matriz de confusión del CNA no dominado de la familia uno, resultado de su operación de clasificación de una muestra de entrenamiento de proporciones homogéneas y cuya cardinalidad total es igual a 165 patrones. Obsérvense las frecuencias relativas de aciertos de clasificación las cuales son 0.9455, 0.9818, 0.9273, para mitosis núcleos y basura, respectivamente. Los tres valores anteriores son valores que se encuentran en los intervalos definidos que conforman la región factible de frecuencias relativas o porcentajes de clasificación que definimos en la ecuación 2.2. La frecuencia relativa total de aciertos de clasificación fue 0.9515, lo que obviamente indica que el porcentaje de clasificaciones erróneas fue menor a 5%, esto indica y como se muestra en la tabla 3.1a, que de un total de 165 patrones sólo ocho patrones fueron mal clasificados de los cuales:

- tres fueron clasificados como mitosis siendo realmente patrones de la clase basura.
- un patrón fue mal clasificado como núcleo cuando ciertamente se trataba de un patrón basura.
- cuatro patrones fueron erróneamente clasificados como pertenecientes a basura pero verdaderamente tres son de la clase mitosis y uno es de la clase núcleos.

Tales errores de clasificaciones dan como resultado los valores de 0.0545, 0.0182 y 0.0727 para los patrones de mitosis, núcleos y basura, respectivamente, que son las frecuencias relativas de falsos positivos que reportó el CNA. Así en detalle, para la muestra de entrenamiento después de evolucionar cien generaciones, un CNA del conjunto de elementos no dominados de la familia uno, clasificó correctamente:

- 52 patrones de 55 patrones pertenecientes a mitosis,
- 54 patrones de la clase núcleos de 55 posibles y
- 51 patrones relacionados a basura de 55 que existían de esa clase en la muestra de entrenamiento.

Estos resultados muestran que el CNA de la figura 3.1, perteneciente a la familia uno, tanto genotípicamente como fenotípicamente fue bien evolucionado y en consecuencia logró encontrar una buena combinación de funciones de activación, pesos y términos de

tendencia que dio buenos resultados de clasificación con la muestra de entrenamiento. Estos resultados eran esperados, ya que estamos hablando de clasificaciones con la muestra de entrenamiento, mas recuérdese cuál es la prueba importante de todo clasificador. Esta se presenta cuando se clasifica la muestra de validación, veamos como se comporta el mismo CNA en la etapa de validación. Analicé la tabla 3.2b.

Veamos en la figura 3.2 la tabla 3.2b, en ella se muestran los resultados de clasificación cuando se presentó la muestra de validación a un CNA no dominado final de la familia uno. Obsérvense las frecuencias relativas de aciertos, 0.89, 0.87, y 0.84 aproximadamente para mitosis, núcleos y basura, respectivamente. Comparando estas frecuencias relativas de clasificaciones correctas con las frecuencias relativas de aciertos de la tabla 3.2, notamos que se presenta un abatimiento aproximadamente del 4%, 10% y 8% para mitosis, núcleos y basura, respectivamente. Mas sin embargo, tales porcentajes de abatimiento no evitaron que en la prueba de validación el CNA no dominado de la figura 3.1 mantuviera sus porcentajes de clasificaciones correctas dentro de la región de clasificaciones factibles que determinamos en la función multiobjetivo de la ecuación 2.2. Este resultado de alguna manera hace que este CNA no dominado tenga capacidad de generalización y por consiguiente sea robusto. Con más detalle, en el último renglón de la tabla 3.2b podemos ver las frecuencias relativas o porcentajes falsos positivos, los cuales aproximadamente son: 28% para mitosis, 4% para núcleos y 13% para basura, en comparación con los resultados con la muestra de entrenamiento, cuyos porcentajes de falsos positivos son aproximadamente 5%, 1% y 7% para mitosis núcleos y basura, respectivamente. Esto indica que aproximadamente para la clase mitosis se cometió un 23% más de errores de clasificación, para la clase núcleos se cometió un 3% más de clasificación incorrectas y para la clase basura se cometió un 6% más de clasificaciones erróneas. Tales incrementos en los porcentajes de falsos positivos dan la impresión de ser muy dramáticos y consecuentemente hacer parecer que el CNA no dominado de la figura 3.1, en su etapa de validación, aproximadamente quintuplica, cuadruplica y duplica sus errores al clasificar mitosis, núcleos y basura, respectivamente. Pero, recuérdese que estamos hablando de frecuencias relativas o porcentajes, y comparar frecuencias relativas sin tomar en cuenta las proporciones induce al error. Así que debemos considerar que la muestra de validación tiene una cardinalidad de aproximadamente cuatro y media veces la cardinalidad de la

muestra de entrenamiento, lo cual nos da una cierta idea de las proporciones. Pero a detalle, las proporciones de las distribuciones por clase también son importantes: la cardinalidad de la clase mitosis en la muestra de validación es aproximadamente poco más de dos veces la cardinalidad de la clase mitosis en la muestra de entrenamiento, la cardinalidad de la clase núcleos en la muestra de validación es aproximadamente cinco veces la cardinalidad que tiene la misma clase en la muestra de entrenamiento y la cardinalidad de la clase basura en la muestra de validación es aproximadamente seis veces la cardinalidad que se observa en la muestra de entrenamiento para la clase basura. Estas proporciones ayudan a comprender los, aparentemente, dramáticos incrementos en los porcentajes de falsos positivos, entre muestra de entrenamiento y muestra de validación. Esto es consecuencia del aumento proporcional en el número de clasificaciones. Como lo muestra la matriz de confusión de la tabla 3.2b, para la muestra de validación el CNA no dominado de la figura 3.1 en detalle reporta:

- 160 patrones clasificados como mitosis de los cuales 114 realmente son mitosis, 4 son originalmente núcleos y 42 son patrones de clase basura.
- 253 patrones son asociados a la clase núcleos, de ellos 242 en efecto son patrones de núcleos y 11 patrones pertenecen a la clase basura.
- relaciona 322 patrones a la clase basura, 279 son en verdad patrones basura, aunque 13 acertadamente son mitosis y 30 verdaderamente son de la clase núcleos.

Los resultados anteriores hacen ver más claramente la relación entre proporciones y, porcentajes o frecuencias relativas, con las que los CNAs deben lidiar. Así que se puede concluir que en la etapa de validación el CNA no dominado de la figura 3.1 clasificó erróneamente cien patrones de un total de 735 patrones clasificados. Esto da como resultado, aproximadamente, un porcentaje total de clasificaciones erróneas del 14% y consecuentemente un porcentaje total de aciertos de clasificación del 86% aproximadamente, como lo muestra el valor de la matriz de confusión en la última columna y último renglón de la tabla 2.2b. Podemos comparar éste valor con ése mismo valor en la matriz de confusión de la tabla 2.2a, cuyo valor da aproximadamente un total de aciertos de clasificación del 95%, lo cual es consecuencia directa porque de un total de 165 patrones

sólo ocho fueron mal clasificados. Así podemos entonces hablar, que entre la muestra de entrenamiento y validación, el porcentaje total de aciertos de clasificación se abate solo un 9% aproximadamente, valor que tomando en cuenta las proporciones entre muestra de entrenamiento y muestra de validación, anteriormente analizadas, reiteradamente nos inducen a concluir que el CNA no dominado, seleccionado aleatoriamente, de la familia uno es un CNA robusto y posee una buena medida de generalidad producto de la metodología evolutiva multiobjetivo y multipoblación que proponemos.

Finalmente de forma comparativa, en ambas tablas de la figura 3.2 se muestra la suma del error cuadrático cuyos valores son 34.13 y 165.069 para la muestra de entrenamiento y para la muestra de validación respectivamente. La diferencia entre ambos valores es bastante significativa, y puede llevar a pensar que en la etapa de validación se cometieron, poco menos de cinco veces, más errores de clasificación que con la muestra de entrenamiento y en resumen concluir que el CNA fue mal entrenado. Pero nuevamente las proporciones son cruciales. Para ver esto regresemos a ver en ambas tablas, además del error cuadrático se muestra el error cuadrático promedio, cuyos valores son aproximadamente 0.207 y 0.225, respectivamente para la muestra de entrenamiento y de validación. Obviamente existe un incremento, pero deja de ser en términos de órdenes de magnitud y se reduce a magnitudes decimales. Estos cambios drásticos en las magnitudes entre error cuadrático y el error cuadrático promedio se entienden por que en la muestra de entrenamiento se están clasificando un total de 165 patrones y en la muestra de validación estamos hablando de un total de 735 patrones. Otra vez las magnitudes son importantes, la muestra de validación tiene una cardinalidad aproximadamente cuatro y media veces la cardinalidad de la muestra de entrenamiento, si nos fijamos que aproximadamente ese mismo número de veces se realizan más errores de clasificación entre la muestra de entrenamiento y la muestra de validación. Estos valores hacen concluir correctamente que el CNA con la muestra de validación, si bien cometió más errores que con la muestra de entrenamiento, lo cual era esperado, el CNA mantiene solidez en su funcionalidad con datos que no le fueron nunca antes presentados. Este hecho hace concluir que el CNA de la figura 3.1 por si solo es un CNA robusto. Mas sin embargo recordemos que nuestra propuesta metodológica tiene una combinación de técnicas no basadas en Pareto con técnicas basadas en Pareto, lo cual de

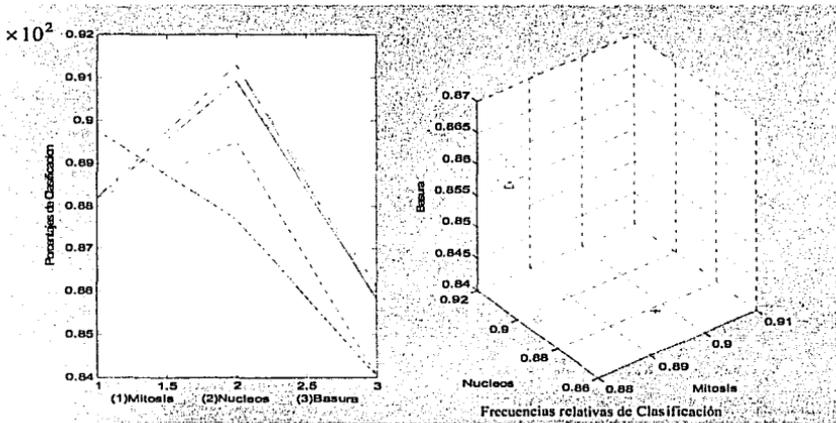
manera conveniente no da una solución única sino un conjunto de soluciones, característica que como antes mencionamos, fortalece más la solución que ofrece nuestra propuesta.

Dada la heurística que establecimos para evaluar el desempeño de los CNAs utilizando ciertos elementos de la matriz de confusión como alternativa de cambio a la expresión matemática del error cuadrático promedio, la función objetivo a minimizar que definimos en la ecuación 2.2, es una función multiobjetivo definida en  $\mathfrak{R}^5$ , cuyos elementos están definidos en base a:

- los valores de los tres porcentajes de aciertos de clasificación, restringidos por tres funciones de membresía determinadas respectivamente en tres intervalos de clasificaciones correctas
- una función de sumatoria determinada por las frecuencias relativas (porcentajes) de falsos positivos y
- una función de varianza cuyos argumentos son las frecuencias relativas (porcentajes) de aciertos de clasificación.

Los cinco elementos anteriormente expuestos, nos ayudan a seleccionar CNAs óptimos, con buenos porcentajes de clasificaciones correctas, con el menor número de errores de clasificación y no sesgados. Así como podemos ver, los elementos de la función multiobjetivo que determinamos, nos dan simultáneamente métricas cualitativas y cuantitativas para evaluar los CNAs como analizamos en el capítulo dos. Pero es obvio que la función multiobjetivo de la ecuación 2.2 tiene efectos directos sobre los porcentajes de aciertos de clasificaciones. Aunado a esto, advertimos que en el análisis comparativo de la dos tablas de la figura 3.2, que contienen respectivamente las dos matrices de confusión de las muestras entrenamiento y validación del CNA no dominado de la figura 3.1, los valores de los porcentajes o frecuencias relativas de aciertos de clasificaciones tuvieron un papel más descriptivo respecto al fenotipo del CNA que las sumatorias y las varianzas asociadas a la ecuación 2.2. Además, debido a que no podemos visualizar gráficas con más de tres dimensiones, en sustitución podemos apreciar de manera directa los efectos de la minimización-maximización de nuestra función multiobjetivo en la sola maximización de los tres porcentajes de aciertos de clasificación de la matriz de confusión, como una función

a maximizar en  $\mathbb{R}^3$ , donde cada objetivo es el porcentaje de aciertos de clasificación de mitosis, núcleos y basura, respectivamente. En la figura 3.3 podemos ver el conjunto óptimo de Pareto final, es decir, el conjunto de CNAs no dominados finales de la familia uno en su etapa de validación.



Gráfica 3.3a

Gráfica 3.3b

Figura 3.3 Conjunto final de CNAs no dominados de la familia uno

Obsérvense las dos gráficas en la figura 3.3, ambas muestran el mismo conjunto de datos pero en forma diferente. En la gráfica 3.3a vemos el conjunto final de CNAs no dominados expresado en coordenadas paralelas, en donde cada objetivo es asociado con un índice cuyo valor entero se sitúa sobre el eje de las abscisas. En la gráfica 3.3b vemos el mismo conjunto final de CNAs no dominados pero en un espacio tridimensional, donde a cada eje del espacio tridimensional le corresponden las frecuencias relativas (porcentajes) de clasificaciones correctas de cada una de las tres clases. El hecho de mostrar ambas gráficas simultáneamente obedece a que la facilidad de estudio que ofrece el método de coordenadas paralelas, se incrementa cuando se combina con las gráficas tridimensionales,

lo cual facilita aún más el análisis ya que nos presentan distintas perspectivas de un misma solución y también nos da la capacidad de comparación entre distintas perspectivas

En ambas gráficas de la figura 3.3 podemos percibir que los efectos de la deriva genética en efecto fueron atenuados, ya que como se observa el conjunto final de CNAs no dominados de la familia uno logró distribuir de forma óptima los porcentajes de aciertos de clasificación en todas las tres alternativas, mitosis núcleos y basura de forma simultánea.

Analicemos con más detalle la gráfica de coordenadas paralelas 3.3a. En ella podemos notar claramente, en el eje de las ordenadas que para las tres clases mitosis, núcleos y basura, los elementos no dominados están entre el 84% y el 92% de porcentajes de aciertos de clasificación. Ambas cotas inferior y superior de porcentajes alcanzados por la familia uno están dentro de los porcentajes de clasificaciones factibles que determinamos en la ecuación 2.2. En la gráfica de coordenadas paralelas 3.3a también podemos observar, en particular un CNA no dominado que se distingue del resto de los CNAs no dominados de la familia, por que tiene el mayor porcentaje de aciertos para mitosis, superior al 89% aproximadamente y el menor porcentaje de aciertos de clasificación para núcleos, 87% aproximadamente; en cuanto a su clasificación de patrones basura es aproximadamente superior al 84%. Este particular CNA se ganó su lugar en el conjunto final de CNAs, por que, además de que sus porcentajes de clasificación de núcleos y basura son superiores al 80%, como lo contempla la ecuación 2.2, singularmente clasifica mejor patrones mitosis que todos los CNAs en la familia uno. Mas ¿qué nos motiva a fijarnos en ese CNA en particular? Aparte de que tal CNA es el mejor clasificando mitosis, el CNA que ha llamado nuestra atención es el mismo CNA de la figura 3.1, cuyos detalles de sus resultados de clasificación con la muestra de entrenamiento y de validación se muestran en las matrices de confusión contenidas en las dos tablas en la figura 3.2. El CNA de la figura 3.1, como ya vimos, es claramente distinguible en la gráfica 3.3a, pero revisando los cromosomas en la familia uno podemos ver que el CNA de la figura 3.1 tiene una réplica exacta, un CNA idéntico tanto genotípicamente como fenotípicamente. Este fenómeno de replicación no se puede apreciar muy bien en la gráfica 3.3a, no obstante que cuidamos en todas las gráficas de coordenadas paralelas asignar a cada línea un tipo diferente de trazo continuo, segmentado, punteado entre otros. Pero debido a que las líneas que unen los porcentajes de

los CNAs no dominados se enciman en la gráfica 3.3a. Este es más evidente en la gráfica tridimensional de la figura 3.3b.

En la gráfica 3.3b podemos ver un conjunto de puntos sobre un espacio tridimensional. Cada punto representa un CNA no dominado de la familia uno, y las tres coordenadas  $(x, y, z)$  asociadas a cada CNA representan los porcentajes o frecuencias relativas de aciertos de clasificación para mitosis núcleos y basura respectivamente. Así en la gráfica 3.3b podemos ver un punto muy definido con coordenadas  $(0.8976, 0.8768, 0.8404)$  aproximadamente. Inmediatamente se deduce que este CNA corresponde con el CNA de la figura 3.1. El hecho de que este punto, asociado al CNA de la figura 3.1, este muy definido sobre la gráfica 3.3b obedece a que en realidad son dos CNAs que están ocupando el mismo punto, ya que a cada punto en las gráficas tridimensionales le asignamos individualmente un símbolo como un rombo, un asterisco, un punto, un triángulo, una cruz, un cuadrado entre otros. Esto hizo percatarnos de la replicación de CNAs en la familia uno, ya que en el mismo punto están sobrepuestos dos símbolos diferentes, y revisando dentro del conjunto de CNAs no dominados de la familia uno comprobamos que en efecto, dos CNAs tienen la misma combinación de funciones de activación, dimensiones de las capas intermedias y obviamente las mismas matrices de pesos y vectores de tendencia. En síntesis, genotípicamente son idénticos y por resultado fenotípicamente también lo son. De hecho en la figura 3.3b podemos notar que el caso de replicación del CNA de la figura 3.1 no es aislado. Obsérvese que en la gráfica 3.3b tenemos una zona de aglomeración de puntos en la región delimitada aproximadamente por los intervalos  $(0.88, 0.89)$ ,  $(0.9, 0.92)$  y  $(0.855, 0.865)$  para mitosis, núcleos y basura, respectivamente. En esa zona de aglomeración aparentemente están colocados cuatro CNAs, que realmente son cinco, pues el punto que está situado en las coordenadas  $(0.881, 0.909, .859)$  aproximadamente, en verdad es un punto ocupado por dos CNAs resultado del fenómeno de replicación. Así en esa zona de aglomeración podemos entonces hablar de cinco CNAs no dominados, los cuales alternan dominancia entre las dimensiones de la región. Quizá por la perspectiva que empleamos para todas las gráficas tridimensionales,  $-37.5^\circ$  de rotación horizontal y  $30^\circ$  de elevación vertical, no sea fácil situar de forma exacta los puntos sobre el espacio que representan las frecuencias relativas alcanzadas por los CNAs en la etapa de validación,

pero esta fue la perspectiva que más se complementa con las gráficas de coordenadas paralelas para el análisis. Ya que si tratamos de buscar, en la gráfica 3.3a, los CNAs de la región de aglomeración de la gráfica 3.3b, la podemos observar mediante un conjunto de líneas que también se aglomeran formando un trazo triangular muy distintivamente y cuyas alternaciones de dominancias se dan en los intervalos  $(0.88, 0.885)$ ,  $(0.909, 0.912)$  y  $(0.859, 0.863)$  aproximadamente para mitosis, núcleos y basura, respectivamente. Más aún, en la gráfica 3.3b podemos notar un punto con coordenadas  $(0.8819, 0.8949, 0.8404)$  aproximadamente. Este CNA en comparación con el CNA de la figura 3.1, clasifica mejor los patrones asociados a núcleos y clasifica igual los patrones basura. Tal CNA lo podemos ver en la figura 3.4.

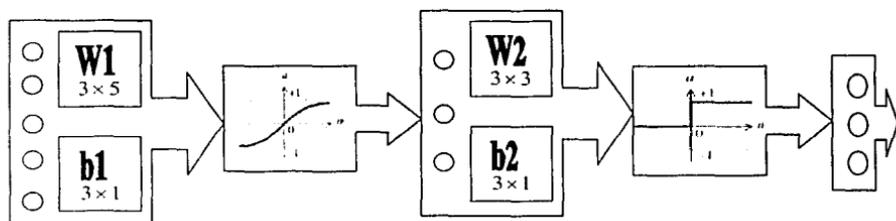


Figura 3.4 Configuración de la arquitectura de un CNA no dominado de la familia uno

Genotípicamente el CNA de la figura 3.4 es diferente del CNA de la figura 3.1. Ambos CNAs tienen distintas combinaciones de funciones de activación, diferente dimensión en la capa intermedia y obviamente las matrices de pesos y los vectores de tendencia son diferentes. En conclusión, genotípicamente son diferentes y sin embargo ambos son CNAs finales y no dominados de la familia uno. A detalle el CNA de la figura 3.4 como se puede ver tiene: tres neuronas en la capa intermedia, y una combinación de funciones de activación tangente hiperbólica y escalón positiva para las capas intermedia y de salida respectivamente a diferencia del CNA de la figura 3.1 cuya configuración es: función de activación escalón simétrica en la capa intermedia, nueve neuronas en la capa intermedia, y en la capa de salida función lineal saturada. Ahora evaluémoslos fenotípicamente. Como se

# TESIS CON FALLA DE ORIGEN

puede apreciar mejor en la gráfica 3.3a de coordenadas paralelas, el CNA de la figura 3.1 clasifica mejor los patrones mitosis que el CNA de la figura 3.4, pero el CNA de la figura 3.4 clasifica mejor núcleos, y en cuanto a patrones asociados a basura, ambos CNAs los clasifican muy cercanamente igual. Así entonces ¿cuál es el mejor de estos dos CNAs?, ya que como lo pide la ecuación 2.2, ambos están dentro de la región de porcentajes aceptables de aciertos de clasificación (ver gráficas en la figura 3.3). Además las frecuencias relativas de aciertos, del CNA de la figura 3.1, tienen una varianza igual a  $8.3824 \times 10^{-4}$  mientras que el valor de la varianza de las frecuencias relativas de aciertos del CNA de la figura 3.4 es igual a  $8.1025 \times 10^{-4}$ . Ambos valores de las varianzas son, además de precisos muy expresivos, ya que nos indican que el espaciamiento entre los valores de las frecuencias relativas de aciertos de clasificación de cada CNA es muy pequeño, lo cual motiva a inducir que ambos CNAs no están sesgados. La suma de los errores de clasificación para el CNA de la figura 3.1 es igual a 0.4646, mientras que la suma de los errores de clasificación para el CNA de la figura 3.4 es 0.3888, como podemos ver en la tabla de la figura 3.5.

Distribución de los patrones		Resultados del Clasificador				
Clases	# de patrones	Mitosis	Núcleos	Basura	# de Errores	Frecuencia relativa de Aciertos
Mitosis	127	112	2	13	15	<b>0.8819</b>
Núcleos	276	2	247	27	29	<b>0.8949</b>
Basura	332	13	40	279	53	<b>0.8404</b>
Total	735	127	289	319	97	<i>0.8680</i>
Frecuencia relativa de Falsos positivos		<b>0.1181</b>	<b>0.1453</b>	<b>0.1254</b>		
Suma del error cuadrático: 196.000000						
Error cuadrático promedio 0.266667						

Figura 3.5 Tabla 3.5 con los valores de la matriz de confusión resultante de la etapa de validación del CNA no dominado de la figura 3.4

De las tablas 3.2b y 3.5, en las figuras 3.2 y 3.5, respectivamente, los dos valores de las sumatorias de las frecuencias relativas o porcentajes de falsos positivos indican por su magnitud que ambos CNAs, en general, cometen pocos errores de clasificaciones. Pero comparando con más detalle, la varianza de las frecuencias relativas de aciertos de clasificación es menor en los resultados del CNA de la figura 3.4 que la varianza de las frecuencias relativas de aciertos de clasificación del CNA de la figura 3.1. Esta mínima

diferencia se nota más en el número total de errores. En la tabla 3.2b podemos ver que el CNA de la figura 3.1 cometió un total de 100 errores, mientras que el CNA de la figura 3.4 cometió 97 errores en total, como comprobamos es muy poca la diferencia. Comparemos más detenidamente las tablas 3.2b y 3.5 de las figuras 3.2 y 3.5, respectivamente.

El CNA de la figura 3.4 clasificó:

- 127 patrones los clasificó como mitosis, siendo efectivamente 112 patrones de mitosis, dos correspondientes a núcleos y 13 pertenecen a patrones de la clase basura.
- 289 patrones los asoció a núcleos, 247 si son patrones núcleos, pero ciertamente dos son patrones de mitosis y 40 son patrones basura.
- 319 patrones los tipificó como basura, de ellos 279 verdaderamente son patrones de basura, 13 realmente son patrones de mitosis y 27 acertadamente son patrones núcleos.

Estos números de patrones mal clasificados tienen una concentración menor a comparación de los resultados del CNA de la figura 3.1. Como podemos ver en la tabla 3.2b, y vale la pena nuevamente mencionar, en su momento el CNA de la figura 3.1 clasificó:

- 160 patrones clasificados como mitosis de los cuales 114 realmente son mitosis, 4 son originalmente núcleos y 42 son patrones de clase basura.
- 253 patrones son asociados a la clase núcleos, de ellos 242 en efecto son patrones de núcleos y 11 patrones pertenecen a la clase basura.
- y relaciona 322 patrones a la clase basura, 279 son en verdad patrones basura, aunque 13 acertadamente son mitosis y 30 verdaderamente son de la clase núcleos.

Nuevamente las proporciones son muy importantes, y nos sitúan en el contexto preciso para establecer una comparación justa. Como podemos ver el CNA de la figura 3.4 cometió menos errores al momento de clasificar mitosis, se equivocó en 15 patrones de 127; mientras que el CNA de la figura 3.1, para mitosis, se equivocó en 46 patrones de 160, a pesar que el CNA de la figura 3.1 acertó 114 de 160 patrones de mitosis y el CNA de la figura 3.4 acertó 112 de 127. Esto hace pensar en la afirmación que existe una sensibilidad de los CNAs hacia la clase mitosis, resultado de su menor representación proporcional. Además como lo afirman los expertos, es muy importante distinguir bien los patrones mitosis de los patrones núcleos y basuras, pero sin dejar de clasificar bien las tres clases.

Con todos los elementos de análisis anteriores, de los dos CNA que hemos estudiado en la familia uno, ¿podemos discernir y declarar únicamente un CNA como el mejor? ¿cuál es, el de la figura 3.1 o el de la figura 3.4? Ya en resumen sabemos que el CNA de la figura 3.4 tiene en sus frecuencias relativas de aciertos de clasificación una varianza menor que el CNA de la figura 3.1, lo cual de alguna forma nos indica que el CNA de la figura 3.4 es menos sesgado. También la suma de las frecuencias relativas de falsos positivos del CNA de la figura 3.4 es menor a la suma de falsos positivos del CNA de la figura 3.1. Pero en los porcentajes o frecuencias relativas de aciertos de clasificación es donde resalta el hecho de que en la etapa de validación, el CNA de la figura 3.1 es quien, dentro de la familia uno, mejor clasifica los patrones de mitosis. Así la alternancia entre los objetivos determinados por la ecuación 2.2 hace que, de entre el CNA de la figura 3.1 y el CNA de la figura 3.4 no nos permita establecer un discernimiento definido y solo nos alcance a poder decir que ambos CNAs son equivalentes para los fines de la función multiobjetivo de la ecuación 2.2: tener buenos porcentajes de aciertos de clasificación, cometer el menor número de errores de clasificación y ser no sesgados.

Todavía más, en la figura 3.3, en particular en la gráfica 3.3a, si enfocamos nuestra atención en las frecuencias relativas de aciertos de clasificación que alcanzó el CNA de la figura 3.4 (0.8819, 0.8949, 0.8404), podemos ver que según las definiciones 1.1 y 1.2 del capítulo uno, el CNA de la figura 3.4 parece estar dominado por el subconjunto de CNAs no dominados que forman una vecindad muy definida mediante una serie de puntos aglomerados en la figura 3.3b, y como ya habíamos anotado, cuyas alternaciones de dominancias se dan en los intervalos (0.88, 0.885), (0.909, 0.912) y (0.859, 0.863)

aproximadamente. ¿Por qué el CNA de la figura 3.4 está dentro del conjunto final de CNAs no dominados, siendo dominado aparentemente por un subconjunto de CNAs no dominados? Nuevamente recordemos que la función multiobjetivo que definimos y que expresamos mediante la ecuación 2.2, es una función vectorial con cinco dimensiones. En ambas gráficas de la figura 3.2 estamos mostrando solo tres de los cinco componentes, es decir, estamos realizando una proyección de nuestra función multiobjetivo definida en  $\mathfrak{R}^5$  al espacio tridimensional  $\mathfrak{R}^3$ , con el fin de poder visualizar y facilitar el análisis. En la ecuación 2.2, los cinco elementos que la componen presentan diferentes escalas de magnitudes, definida la ecuación 2.2 como:

- Maximizar las frecuencias relativas (los porcentajes) de aciertos de clasificaciones de las tres clases de patrones y restringirlos a que se mantengan en el intervalo  $[0.8, 0.99]$ .
- Minimizar los valores de las varianzas de las frecuencias relativas de los CNAs no dominados y como ya hemos visto, para los CNAs no dominados anteriores, tienen una magnitud alrededor de  $10^{-4}$ .
- Minimizar la magnitud de la sumatoria de las frecuencias relativas de falsos positivos y como también ya comprobamos, los valores óptimos de este elemento están alrededor de 0.4.

La forma de pasar de un problema de maximización a un problema de minimización involucra una sencilla operación lineal, de hecho la aplicamos en la programación que realizamos. También sabemos que podemos evitar el problema de la diferencia de escalas mediante una operación de normalización a todos los componentes de la ecuación 2.2 en un intervalo determinado. Mas ambas operaciones de una forma u otra no nos permitirían apreciar directamente los porcentajes de aciertos de clasificación, o sacrificaríamos la magnitud real de la suma de las frecuencias relativas de falsos positivos, o perderíamos la magnitud real de la varianza que alcanzan las frecuencias relativas de aciertos de los CNAs. Y de los cinco componentes de la ecuación 2.2, los tres porcentajes ( frecuencias relativas )

de aciertos de clasificaciones son los componentes más indicativos y expresivos del desempeño de los CNAs, además de que los tres porcentajes de aciertos de clasificaciones están en la misma escala de magnitudes, en consecuencia, de la presión que imprime en la selección, la suma de los falsos positivos y la varianza en la selección de los CNAs no dominados. Con esta justificación que apoya a la realizada al inicio de este capítulo, al hecho de sólo graficar los porcentajes de aciertos de clasificación, podemos explicar por que el CNA de la figura 3.4 aparece en la gráfica 3.3a como perteneciente al conjunto de CNAs no dominados, siendo aparentemente dominado. La explicación radica en los elementos que mostramos en las gráficas 3.3a y 3.3b, los porcentajes de aciertos de clasificaciones correctas, si bien son los indicadores más expresivos del desempeño de los CNAs, pero en particular el CNA de la figura 3.4, tuvo una distribución más homogénea de errores de clasificación. Esto se muestra en la síntesis de las frecuencias relativas de falsos positivos en la tabla de la figura 3.5 y cuyos valores son: 0.1181, 0.1453 y 0.1254 para mitosis, núcleos y basura, respectivamente, a diferencia de las frecuencias relativas de falsos positivos del CNA de la figura 3.1 cuyos valores son: 0.2875, 0.0435 y 0.1335 para mitosis, núcleos y basura, respectivamente. Como vemos el CNA de la figura 3.4 es quien logró de forma más homogénea distribuir los errores de clasificación en su etapa de validación, pero no exclusivamente en referencia al CNA de la figura 3.1, sino que dentro de la familia uno, el CNA de la figura 3.4 fue el mejor en ese aspecto. Para verificar este hecho, tomemos un CNA del subconjunto de CNAs no dominados de la familia uno, que anteriormente habíamos destacado porque, en la gráfica 3.3b, forman muy marcadamente una serie de puntos aglomerados y también, muy distintivamente, describen una serie de líneas acumuladas en la gráfica 3.3a, que forman un trazo triangular muy definido. El CNA de este subconjunto lo mostramos en la figura 3.6, como podemos analizar, genotípicamente tiene una configuración distinta de los CNAs de las figuras 3.1 y 3.4. El CNA de la figura 3.6 tiene: en la capa intermedia una función de activación lineal, seis neuronas en la capa intermedia y en la capa de salida una función de activación sigmoide logarítmica, además de obviamente cinco neuronas en la capa de entrada y tres neuronas en la capa de salida.

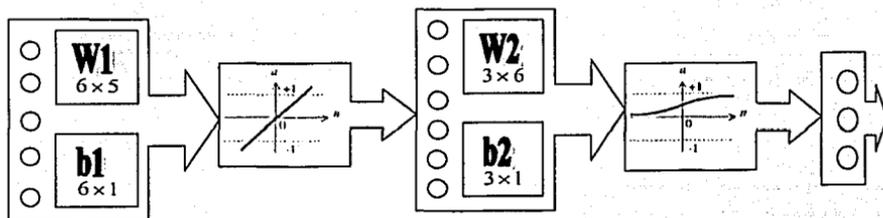


Figura 3.6 Configuración de la arquitectura de un CNA no dominado de la familia uno

Pero con solo ver un elemento de la familia no podemos inferir nada relacionado con el genotipo del resto de los CNAs en la familia. Así, si del subconjunto de CNAs no dominados que forman, de manera muy distinguible una serie de puntos o líneas aglomeradas en ambas gráficas de la figura 3.3, tomamos un CNA y que tal CNA tenga el genotipo de la figura 3.6, no indica que todos los CNAs de ese subconjunto compartan esa configuración genotípica. Si acaso por la característica de formar una aglomeración alrededor de una vecindad en determinados porcentajes de aciertos de clasificación, un análisis fenotípico de un CNA del subconjunto proporcione ideas aproximadas del comportamiento del resto de los CNAs en el subconjunto, son solo eso, aproximaciones. Veamos cual es el fenotipo del CNA de la figura 3.6. Esto es, veamos su matriz de confusión en la etapa de validación ( figura 3.7 ).

Distribución de los patrones		Resultados del Clasificador				
Clases	# de patrones	Mitosis	Núcleos	Basura	# de Errores	Frecuencia relativa de Aciertos
Mitosis	127	112	2	13	15	0.8819
Núcleos	276	23	252	1	24	0.9130
Basura	332	42	5	285	47	0.8584
Total	735	177	259	299	86	0.8830
Frecuencia relativa de Falsos positivos		0.3672	0.0270	0.0468		
Suma del error cuadrático: 104.899998						
Error cuadrático promedio: 0.142721						

Figura 3.7 Tabla 3.7 con los valores de la matriz de confusión resultante de la etapa de validación del CNA no dominado de la figura 3.6

TESIS CON  
FALLA DE ORIGEN

De la comparación directa entre la tabla en la figura 3.7 y la tabla en la figura 3.5 vemos que la suma de las frecuencias relativas de falsos positivos para el CNA de la figura 3.4 es igual a 0.3888 y la suma de las frecuencias relativas de falsos positivos del CNA de la figura 3.6 es 0.441. Pero, la varianza de las frecuencias relativas de aciertos de clasificación del CNA de la figura 3.6 es igual a  $7.5010 \times 10^{-4}$ . De inmediato surge el cuestionamiento ¿por qué la suma de las frecuencias relativas de falsos positivos es menor para el CNA de la figura 3.4 en comparación con el CNA de la figura 3.6, si el CNA de la figura 3.6 cometió menos errores y su sesgo es menor, según el valor de su varianza? La explicación radica en el hecho de que, en efecto el CNA de la figura 3.4 cometió 97 errores de clasificación a diferencia del CNA de la figura 3.6 quien tuvo 86 errores de clasificación, pero los errores de clasificación del CNA de la figura 3.4 tuvieron una distribución más uniforme proporcionalmente en comparación con el CNA de la figura 3.6. Veamos las frecuencias relativas de falsos positivos para el CNA de la figura 3.4 son: 0.1181, 0.1453 y 0.1254 para mitosis núcleos y basura, respectivamente; las frecuencias relativas de falsos positivos para el CNA de la figura 3.6 son: 0.3672, 0.0270 y 0.0468 para mitosis núcleos y basura, respectivamente. Esto provoca ver nuevamente la importancia de las proporciones en las muestra de patrones. Podemos ver en la matriz de confusión de la tabla en la figura 3.7 que el CNA de la figura 3.6 clasificó 177 patrones como mitosis, pero acumuló 65 errores ya que acertadamente sólo 112 patrones son mitosis. Como en la muestra de validación tenemos 127 patrones de la clase mitosis, entonces resulta que en la clasificación de los patrones mitosis, el CNA de la figura 3.6, se está equivocando aproximadamente en un 36% y está acertando en un 88%, como podemos ver en la tabla de la figura 3.7. A diferencia del CNA de la figura 3.4 quien para la clase mitosis está acertando en un 88%, pero se está equivocando en un 11% aproximadamente como analizamos en la tabla de la figura 3.5. Nuevamente al ser la clase mitosis la que tiene la menor representación proporcional en la muestra de validación, hace que el CNA sea muy sensible al número de errores en la clase mitosis. Este hecho es importante porque, como lo refieren los expertos, este sesgo hacia la desproporción en la distribución de ciertas etapas de desarrollo celular se presenta tanto en la actual población experimental de patrones de etapas de desarrollo celular, como en las poblaciones naturales de células animales y vegetales. Es una característica propia de la naturaleza del desarrollo celular.

Pero como vemos, al menos para la familia uno, la metodología evolutiva de diseño y entrenamiento que proponemos encuentra CNAs cuyos porcentajes, o frecuencias relativas, en la etapa de validación se abaten muy poco, sin lograr sacar a los porcentajes de aciertos de los CNAs de la región de porcentajes aceptables, no obstante que la muestra de entrenamiento no tiene las desproporciones propias de la población experimental de patrones de desarrollo.

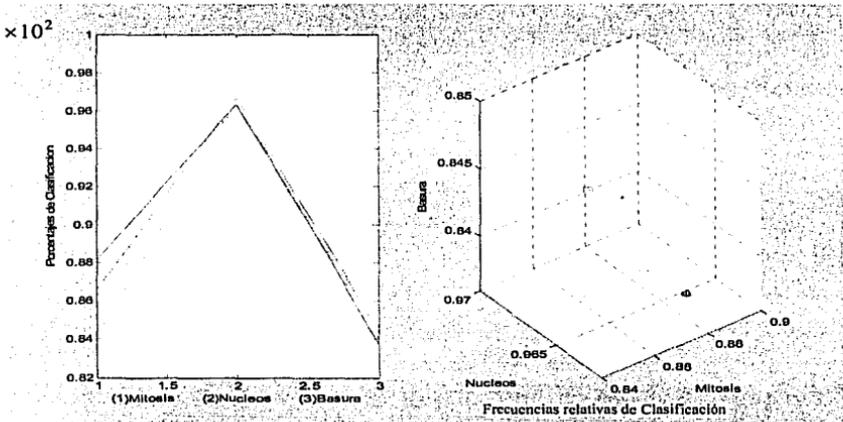
Finalmente, en la figura 3.3 se muestran los resultados de todos los CNAs no dominados de la familia uno. La mayoría genotípicamente son completamente diferentes, y como ya vimos a detalle, las alternaciones de las dominancias en las dimensiones de los valores de la función multiobjetivo no nos permiten discernir y distinguir a un CNA como el mejor, lo cual nos lleva a concluir, con más argumentos, que fenotípicamente son equivalentes.

Si en lo individual, para el conjunto de CNAs no dominados de la familia uno, hablar de un CNA no dominado fenotípicamente es equivalente a hablar de otro CNA no dominado, entonces veamos los resultados en conjunto. Así la cardinalidad del conjunto no dominado de la familia uno es de ocho CNAs, lo cual concluye que para la familia uno, el porcentaje final de CNAs no dominados es del 16% es decir, al final del plan evolutivo ocho de cincuenta CNAs en la familia uno son óptimos, con buenos porcentajes de aciertos de clasificación, con el menor número de errores de clasificación y sin sesgo de clasificaciones. Esto presenta una ventaja en comparación al procedimiento clásico de diseño y entrenamiento de CNAs, en donde el experto en Redes Neuronales regularmente después de un largo proceso de diseño, entrenamiento y validación obtiene únicamente un clasificador óptimo.

Analicemos ahora los resultados de la familia dos de CNAs. En la figura 3.8 podemos ver, en ambas gráficas, el conjunto final de CNAs no dominado de la familia dos en la etapa de validación.

No obstante que durante las generaciones del plan evolutivo, los CNAs de la familia dos se calificaban con la muestra de entrenamiento y se demandaba en la función multiobjetivo que todos los CNAs con mayor valor de aptitud tuvieran frecuencias relativas de aciertos de clasificación dentro del intervalo  $[0.85, 0.99]$ . En la gráfica 3.8a de coordenadas paralelas podemos ver que los porcentajes, o frecuencias relativas, de aciertos de clasificación de los CNAs no dominados finales para mitosis están en el intervalo  $(0.84, 0.885)$ , para núcleos en

el intervalo (0.96,0.97) y para basura en el intervalo (0.835,0.845) aproximadamente. Esto nos indica que aun con una muestra con la cual no fueron evolucionados, los CNAs no dominados finales cumplen con el objetivo de tener los porcentajes de aciertos de clasificación dentro de los intervalos óptimos de clasificaciones.



Gráfica 3.8a

Gráfica 3.8b

Figura 3.8 Conjunto final de CNAs no dominados de la familia dos.

En la gráfica 3.8a es difícil distinguir cual es la cardinalidad del conjunto de CNAs no dominados de la familia dos, debido a que es muy estrecha la alternancia de la dominancia de los porcentajes de aciertos de clasificación de los CNAs no dominados de la familia dos. Veamos la gráfica 3.8b en la figura 3.8.

En la gráfica tridimensional 3.8b podemos apreciar cinco puntos, como ya habíamos mencionado. Cada punto representa un CNA no dominado de la familia dos, y las tres coordenadas  $(x,y,z)$  asociadas a cada CNA representan las frecuencias relativas de aciertos de clasificación para mitosis núcleos y basura, respectivamente. En la gráfica 3.8b podemos distinguir cinco símbolos diferentes asociados a cada punto en el espacio

tridimensional. Si nos fijamos atentamente en el símbolo que esta en el punto con coordenadas aproximadas  $(0.883, 0.965, 0.97)$ , podemos apreciar que son dos símbolos que están ocupando el mismo punto, es decir existe replicación de CNAs no dominados. Recordemos que un caso similar de replicación se había presentado en la familia uno de CNAs. Así entonces el conjunto de CNAs no dominados de la familia dos tiene verdaderamente una cardinalidad de seis elementos.

La cardinalidad del conjunto de elementos no dominados finales de la familia dos es de seis CNAs no dominados. Así podemos decir que en la familia dos el porcentaje de eficacia de la búsqueda de CNAs óptimos es del 12% aproximadamente. Nuevamente y por la mismas razones antes expuestas para la familia uno, esta cantidad de CNAs no dominados encontrados en la familia dos representa por si sola una ventaja frente a los procedimientos convencionales de diseño, entrenamiento y validación de CNAs.

Seleccionemos aleatoriamente un elemento del conjunto final de CNAs no dominados de la familia dos. Este se muestra en la figura 3.9. Como podemos analizarlo genótipicamente tiene: una función de activación lineal positiva en la capa intermedia, ocho neuronas en la capa intermedia y en la capa de salida una función de activación lineal saturada. Esta configuración en la arquitectura del CNA elegido es diferente a las configuraciones de las arquitecturas de los CNAs seleccionados pertenecientes a la familia uno, aunque la función de activación lineal saturada en la capa de salida es un factor que ya se había presentado en el CNA de la figura 3.1 que pertenece a la familia uno. Pero este hecho no es determinante para establecer una relación entre este CNA de la familia dos y el CNA de la familia uno.

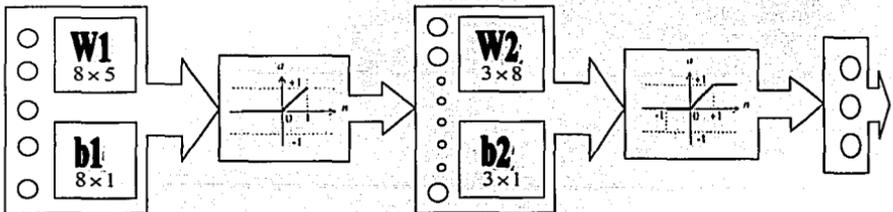


Figura 3.9 Configuración de la arquitectura de un CNA no dominado de la familia dos

El genotipo de un CNA no nos refiere mucho en cuanto a su funcionalidad. Para evaluar el desempeño del CNA de la figura 3.9 y su capacidad de generalización veamos su fenotipo con la muestra de entrenamiento y con la muestra de validación. En la figura 3.10 podemos analizar las dos matrices de confusión del CNA de la figura 3.9, con la muestra de entrenamiento y en la etapa de validación.

Distribución de los patrones		Resultados del clasificador				
Clases	# de patrones	Mitosis	Núcleos	Basura	# de Errores	Frecuencia relativa de Aciertos
Mitosis	55	53	0	2	2	<b>0.9636</b>
Núcleos	55	0	54	1	1	<b>0.9818</b>
Basura	55	0	5	50	5	<b>0.9091</b>
Total	165	53	59	53	8	<i>0.9515</i>
Frecuencia relativa de Falsos positivos		<b>0.0000</b>	<b>0.0847</b>	<b>0.0566</b>		
Suma del error cuadrático: <i>11.896300</i>						
Error cuadrático promedio: <i>0.072097</i>						

Tabla 3.10a Matriz de confusión resultado de la clasificación con la muestra de entrenamiento

Distribución de los patrones		Resultados del clasificador				
Clases	# de patrones	Mitosis	Núcleos	Basura	# de Errores	Frecuencia relativa de Aciertos
Mitosis	127	110	0	17	17	<b>0.8661</b>
Núcleos	276	3	266	7	10	<b>0.9638</b>
Basura	332	27	26	279	53	<b>0.8404</b>
Total	735	140	292	303	80	<i>0.8912</i>
Frecuencia relativa de Falsos positivos		<b>0.2143</b>	<b>0.0890</b>	<b>0.0792</b>		
Suma del error cuadrático: <i>94.712363</i>						
Error cuadrático promedio: <i>0.128860</i>						

Tabla 3.10b Matriz de confusión resultado de la clasificación con la muestra de validación

Figura 3.10 Resultados de clasificación del CNA no dominado de la figura 3.9

En la tabla 3.10a mostramos la matriz de confusión resultado con la muestra de entrenamiento y la tabla 3.10b contiene la matriz de confusión resultado con la muestra de validación. Comparando entre los valores de ambas tablas, podemos ver que los porcentajes de aciertos de clasificación con la muestra de validación en referencia con la muestra de entrenamiento presentaron un abatimiento aproximadamente del 10%, 2% y del 6% para mitosis, núcleos y basura, respectivamente. El pronóstico de abatimiento era esperado,

considerando que los patrones de etapas de desarrollo celular de la muestra de validación no se presentaron durante la evolución de los CNAs de la familia dos. Recordemos que para el CNA de la figura 3.1 los porcentajes de abatimiento entre muestra de entrenamiento y muestra de validación, para mitosis, núcleos y basura fueron aproximadamente 4%, 10% y 8%, respectivamente. Como vemos ningún porcentaje de aciertos de clasificación se abate más allá del 10%. Este hecho es una característica de todos los CNAs no dominados finales de las tres familias y además es una característica que se presentó en los 260 ensayos exitosos de las 300 pruebas que realizamos de nuestra metodología propuesta.

Con mas detalle podemos comprobar, para la etapa de validación, que el valor de la varianza de las frecuencias relativas de aciertos de clasificación del CNA de la figura 3.9 es igual a  $4.2 \times 10^{-3}$  y la suma de las frecuencias relativas de falsos positivos es aproximadamente 0.3825. Obsérvese que el valor de la varianza de las frecuencias relativas de aciertos de clasificación del CNA de la figura 3.9 es alto, en comparación con las varianzas de los CNAs analizados pertenecientes a la familia uno. Esto se explica porque el porcentaje de aciertos de clasificación para la clase núcleos es superior en un valor alrededor del 11% aproximadamente en comparación con los porcentajes de aciertos de clasificación de mitosis y basura, lo cual no se había presentado en los CNAs analizados de la familia uno. Mas no obstante que el valor de la varianza del CNA de la figura 3.9 es alta en comparación con los CNAs de la familia uno, el CNA de la figura 3.9 no es sesgado, ya que las tres frecuencias relativas de aciertos de clasificaciones las cuales son 0.8661, 0.9638, y 0.8404 para mitosis núcleos y basura, respectivamente están dentro de la zona aceptable de frecuencias relativas (porcentajes) de clasificaciones correctas. Es interesante notar que, mientras la varianza de las frecuencias relativas de clasificaciones correctas del CNA de la figura 3.9 es alta en comparación con los CNAs analizados de la familia uno, el CNA de la figura 3.9 tiene un valor de la suma de las frecuencias relativas de falsos positivos menor en comparación con los CNAs de la familia uno. Esto se explica por la razón que el CNA de la figura 3.9 cometió 80 errores, como se nota en la tabla 3.10. Este número de errores es menor en comparación con los CNAs de la familia uno. Además si bien la distribución de los errores del CNA de la figura 3.9 no es tan homogénea proporcionalmente como la distribución de los errores del CNA de la figura 3.5, la

TESIS CON  
FALLA DE ORIGEN

distribución de los errores de clasificación del CNA de la figura 3.9 es homogénea absolutamente por que como podemos extraer de la tabla 3.10b en la figura 3.10:

- 140 patrones fueron clasificados como mitosis, pero 30 están mal clasificados por que realmente pertenecen a núcleos o basura
- 292 patrones fueron clasificados como núcleos, pero 26 verdaderamente son patrones de mitosis o basura
- 303 patrones fueron clasificados como basura, pero acertadamente 24 son patrones asociados a mitosis o núcleos

Esta distribución absoluta de los errores de clasificación, además del porcentaje total de aciertos de clasificaciones correctas aproximadamente 89%, más alto que los porcentajes totales de aciertos de clasificaciones correctas de los CNAs de la familia uno, hacen que el CNA de la figura 3.9 sea un CNA óptimo con buenos porcentajes de aciertos de clasificaciones correctas, con bajos porcentajes de errores de clasificación y no tenga sesgo hacia una u otra clasificación.

Regresemos a analizar globalmente el conjunto de CNAs no dominados de la familia dos. Como podemos ver en la gráfica 3.8a de la figura 3.8, las líneas que unen las coordenadas paralelas están aglomeradas. Esto es porque los CNAs de la familia dos alternan muy cercanamente las dominancias de los porcentajes de aciertos de clasificaciones. En la gráfica 3.8b se amplía la panorámica de los porcentajes de aciertos de clasificación, pero debido a la perspectiva aérea que mostramos, se hace difícil localizar exactamente en el espacio los puntos que corresponden con CNAs de la familia dos y obviamente las comparaciones no son tan sencillas desde este nivel. Pero, recordemos que está es la mejor perspectiva que conjuntamente con las gráficas de coordenadas paralelas nos facilitan el análisis de las familias de CNAs. Del análisis conjunto de ambas gráficas en la figura 3.9, podemos ver que la longitud de los intervalos de las frecuencias relativas de aciertos de clasificaciones correctas, donde se da la alternancia en la dominancia de los CNAs no dominados del la familia dos, son aproximadamente del 0.045, 0.005 y 0.002 para mitosis

núcleos y basura, respectivamente. Esto indica que los CNAs dominantes finales de la familia dos, en conjunto, fenotípicamente son muy similares al CNA de la figura 3.9. En consecuencia las matrices de confusión, en la etapa de validación, para los CNAs finales no dominados de la familia dos son muy similares a la matriz de confusión contenida en la tabla 3.10b en la figura 3.10 y que solo presentan alguna leves variaciones en las frecuencias relativas de aciertos de clasificación. Pero no podemos decir nada referente a la configuración genotípica del resto de los CNAs no dominados restantes de la familia dos. No podemos decir que son muy similares genotípicamente al CNA de la figura 3.8, así que solo nos quedamos con que fenotípicamente son equivalentes y en consecuencia son CNAs con buenos porcentajes (frecuencias relativas) de clasificaciones correctas, con bajos porcentajes de clasificaciones erróneas y sin sesgo.

Revisemos ahora el comportamiento del conjunto final de CNAs no dominados de la familia tres. Para esto, analicemos las dos gráficas en la figura 3.11.

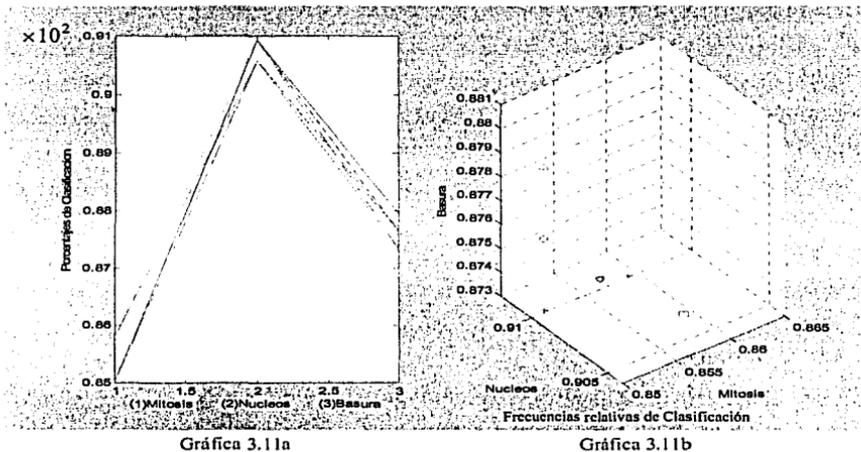


Figura 3.11 Conjunto final de CNAs no dominados de la familia tres.

En la gráfica 3.11a de coordenadas paralelas podemos ver que los intervalos de aciertos de clasificaciones que alcanzaron los CNAs no dominados finales de la familia tres son

aproximadamente:  $[0.85, 0.86]$  para mitosis,  $[0.9053, 0.91]$  para núcleos y  $[0.873, 0.88]$  para basura. Estos intervalos también se pueden apreciar en la gráfica tridimensional 3.11b. Tales intervalos en donde caen las frecuencias relativas de aciertos de clasificación de los CNAs no dominados de la familia tres, implican que se trata de CNAs con porcentajes de aciertos de clasificación que cumplen con lo que demanda la función multiobjetivo que definimos en la ecuación 2.2. Como hicimos para la familia dos, también podemos extraer información adicional de los intervalos anteriores, su longitud nos dice qué tan cercanos están los porcentajes de aciertos de clasificaciones. Entonces, para los CNAs no dominados finales de la familia tres, la longitud del intervalo de clasificaciones correctas de mitosis es igual a 0.01, para núcleos es igual a 0.0047 y para basura es igual a 0.007. Esto implica deducir que la alternancia en las dominancias de las diferentes dimensiones de la función objetivo son muy cercanas en cuanto a frecuencias relativas de aciertos de clasificación, como se nota en la gráfica 3.11b.

Ahora veamos cuantos CNAs no dominados finales en la familia tres alcanzó a diseñar y ajustar nuestra propuesta de diseño y evolución de CNAs. En la gráfica 3.11b es sencillo distinguir seis puntos en el espacio tridimensional, recordemos que cada punto está asociado a un CNA no dominado de la familia tres en base a sus tres porcentajes de aciertos de clasificación. A su vez, cada punto tiene asociado un símbolo diferente para poder distinguir claramente. En la gráfica 3.11b enfoquemos nuestra atención en el punto con coordenadas aproximadas  $(0.8553, 0.91, 0.873)$  para mitosis, núcleos y basura, respectivamente. Si nos fijamos son dos símbolos que ocupan el mismo punto en el espacio, esto es por que como vimos en las dos familias anteriores, se presentó la replicación de CNAs. Revisando a detalle el conjunto de CNAs no dominados de la familia tres, encontramos que su cardinalidad es de siete elementos y que precisamente existían dos CNAs idénticos en fenotipo y genotipo.

Dado que como hemos visto, la longitud de los intervalos de los porcentajes de aciertos de clasificación es muy estrecha, indicio que fenotípicamente son equivalentes, para darnos una idea de cómo son fenotípicamente los CNAs finales no dominados en la etapa de validación de la familia tres, seleccionemos aleatoriamente un CNA no dominado de la familia tres. Véase figura 3.12.

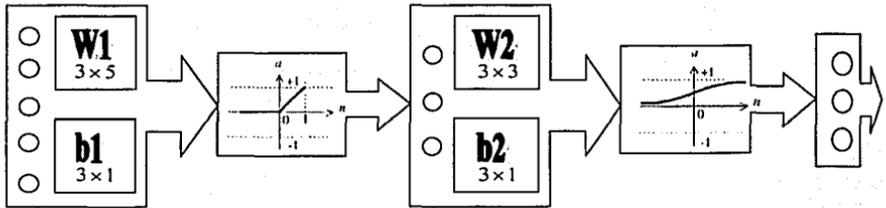


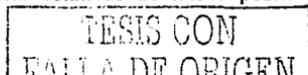
Figura 3.12 Configuración de la arquitectura de un CNA no dominado de la familia tres

La estructura genotípica del CNA seleccionado de la familia tres es: una función de activación en la capa intermedia del tipo lineal positiva, tres neuronas en la capa intermedia y una función de activación sigmoide logarítmica en la capa de salida, como se puede ver en la figura 3.12. Sabemos que el genotipo de un CNA no dominado no indica nada respecto al genotipo del resto de los CNAs no dominados. Veamos en la figura 3.13 su fenotipo, el cual si nos ofrece una idea más aproximada del desempeño del resto de los CNAs no dominados de la familia tres.

Distribución de los patrones		Resultados del Clasificador				
Clases	# de patrones	Mitosis	Núcleos	Basura	# de Errores	Frecuencia relativa de Aciertos
Mitosis	127	108	0	19	19	<b>0.8504</b>
Núcleos	276	3	251	22	25	<b>0.9094</b>
Basura	332	26	15	291	41	<b>0.8765</b>
Total	735	137	266	332	85	<b>0.8844</b>
Frecuencia relativa de Falsos positivos		<b>0.2117</b>	<b>0.0564</b>	<b>0.1235</b>		
		Suma del error cuadrático: 147.416467				
		Error cuadrático promedio: 0.200567				

Figura 3.13 Tabla 3.13 con los valores de la matriz de confusión resultante de la etapa de validación del CNA no dominado de la figura 3.12

De los valores en la matriz de confusión contenida en la tabla 3.13 podemos comprobar que el valor de la varianza de las frecuencias relativas de aciertos de clasificaciones es igual a  $8.7410 \times 10^{-4}$  y el valor de la suma de las frecuencias relativas de falsos positivos es



aproximadamente igual a 0.3916. Ambos valores nos indican que el CNA de la figura 3.12 es un CNA no sesgado, con pocos errores de clasificación y además como ya hablamos notado, sus porcentajes de aciertos de clasificación en la etapa de validación, están dentro de los intervalos que definimos en la ecuación 2.2 y que nos determinan la región de porcentajes aceptables de clasificaciones correctas.

Número de Figura del CNA de referencia	Cantidad Total de errores de clasificación	Suma del Error Cuadrático	Error Cuadrático Promedio	Varianza de las frecuencias relativas de Aciertos de clasificación	Suma de las frecuencias relativas de Falsos positivos
Figura 3.1	100	165.069	0.224584	$8.3824 \times 10^{-4}$	0.4646
Figura 3.4	97	196.0	0.266667	$8.1025 \times 10^{-4}$	0.3888
Figura 3.6	86	104.89998	0.142721	$7.5010 \times 10^{-4}$	0.441
Figura 3.9	80	94.712363	0.128860	$4.2 \times 10^{-3}$	0.3825
Figura 3.12	85	147.416467	0.200567	$8.7410 \times 10^{-4}$	0.3916

Figura 3.14 Tabla 3.14 Resumen de los valores obtenidos en la etapa de validación por los CNAs analizados anteriormente

En síntesis, pero ahora en conjunto revisemos todas las tablas mostradas anteriormente, en particular las de la etapa de validación. Sabemos que ellas contienen las matrices de confusión resultado de la clasificación con la muestra de validación, en todas ellas incluimos los valores del error cuadrático y del error cuadrático promedio. Estos valores junto con la cantidad total de errores de clasificación, los valores de la varianza de las frecuencias relativas de aciertos de clasificación y la suma de las frecuencias relativas de falsos positivos se muestran en la tabla de la figura 3.14. Observando la tabla podemos comprobar que el error cuadrático promedio no es una buena métrica para evaluar el desempeño de CNAs [KUR00a], [KUR00b]. Si establecemos una jerarquía de los CNAs analizados anteriormente, en base al valor de su correspondiente error cuadrático promedio, según podemos extraer de la información en la tabla de la figura 3.14, el CNA de la figura 3.4 tendría la menor jerarquía y el CNA de la figura 3.9 tendría la mayor. Esto de alguna forma se explica por el hecho que el CNA de la figura 3.9 cometió 80 errores mientras que el CNA de la figura 3.4 cometió 97 errores de clasificación. El CNA que cometa más errores de clasificación tiene un desempeño inferior frente al CNA que cometa menos

errores de clasificación. Revisando la tabla en la figura 3.14, podemos ver que el CNA de la figura 3.1 cometió tres errores más que el CNA de la figura 3.4 pero el error cuadrático promedio del CNA de la figura 3.1 tiene un error cuadrático promedio menor que el CNA de la figura 3.4, ¿por qué? Bueno, recordemos cómo se obtiene el error cuadrático: es la suma de las diferencias al cuadrado de las respuestas de de las neuronas de salida y las respuestas deseadas. En el caso de la clasificación de patrones de estados de desarrollo celular, como en su gran mayoría, las salidas deseadas se determinaron como vectores en el espacio de Hamming cuyos elementos son valores binarios, mientras las respuestas de las neuronas de salida dependen de la función de activación de la capa de salida, de los pesos y términos de tendencia asociados a las neuronas de salida. Los términos de tendencia, según su magnitud y signo, trasladan a la función de activación y los pesos según su magnitud y signo controlan la pendiente y dirección respectivamente de la función de activación. Así de la forma más inmediata, pero no completa, la respuesta de las neuronas de salida depende de cual sea la función de activación en la capa de salida. Así las respuestas de las neuronas de salida son vectores  $\mathfrak{R}^n$  cuyos elementos son valores en  $\mathfrak{R}$ , salvo cuando la función de activación de la capa de salida es una función escalón positiva o escalón simétrica positiva, cuyas imágenes son valores binarios o valores  $-1$  y  $1$ , respectivamente. Así, mientras las salidas deseadas son vectores de valores binarios, las salidas del CNA son vectores de valores binarios o vectores de valores reales, dependiendo de la función de activación de la capa de salida. Y como todas las neuronas de salida participan en la respuesta del CNA, entonces no existe una relación lineal entre la cantidad total de errores de clasificación y el error cuadrático o el error cuadrático promedio. Este hecho es evidente en la tabla de la figura 3.14, ya que el CNA de la figura 3.6 cometió 86 errores de clasificación, mientras que el CNA de la figura 3.12 cometió 85 errores de clasificación, lo cual nos haría pensar que el CNA de la figura 3.12 tiene un error cuadrático medio menor que el CNA de la figura 3.6, pero como lo resume la tabla en la figura 3.14 esto es equivocado. Mas aún, en la tabla 3.14 podemos ver que coincide en que el CNA de la figura 3.9 quién cometió menos errores de clasificación, tiene el menor error cuadrático promedio. También coincide con que fue quién tuvo el menor valor de la suma de frecuencias relativas de falsos positivos, pero ese mismo CNA fue quién tuvo el valor más alto de la varianza de las frecuencias relativas de aciertos de clasificación. Tales

alternancias en las jerarquías de los valores para seleccionar los mejores CNAs nos corroboran que una sola medida para establecer la dominancia entre CNAs no funciona, los criterios para evaluar el desempeño de los CNA, en particular deben ser compuestos y complejos que incluyan métricas cuantitativas y cualitativas, pero sobre todo estadísticas.

### **3.2.2 DESARROLLO Y EVOLUCIÓN CON UNA MUESTRA HETEROGÉNEA**

En esta sección vamos a mostrar de forma resumida, algunos resultados de la metodología usada, cuando se seleccionaron muestras de entrenamiento no homogéneas en las distribuciones de las proporciones de cada clase de patrones de etapas de desarrollo celular, pero siempre cuidando que sean representativas de la población de patrones de etapas de desarrollo celular.

Entonces se seleccionó una muestra de entrenamiento con las siguientes proporciones:

- 67 patrones clasificados como mitosis
- 99 patrones asociados a mitosis
- 116 patrones tipificados como basura

La forma en que se seleccionó la muestra de entrenamiento anterior fue mediante una distribución uniformemente aleatoria exigiendo que los valores de las medias de cada parámetro de la muestra, cumplieran con que estén contenidos dentro de los intervalos típicos muestrales que establecimos en el capítulo dos para cada parámetro. Nótese que la cardinalidad de la muestra no homogénea de entrenamiento es igual a 282 patrones, la cual es superior a la muestra homogénea de entrenamiento de la sección 3.2.1 cuya cardinalidad era de 165 patrones. Las proporciones en la muestra de entrenamiento de proporciones heterogéneas, fueron tomadas de tal manera que tuvieran la misma tendencia de las proporciones de la población experimental de patrones de etapas de desarrollo celular y poder analizar si la metodología de diseño y entrenamiento evolutivo y multiobjetivo que proponemos es sensible a las proporciones de la muestra de entrenamiento.

Con la muestra heterogénea de entrenamiento representativa, se ejecutó el algoritmo genético modificado que proponemos, con los mismos parámetros genéticos que mencionamos cuando se seleccionó una muestra homogénea de entrenamiento. El conjunto final de CNAs no dominados de la familia uno se muestra en la figura 3.15.

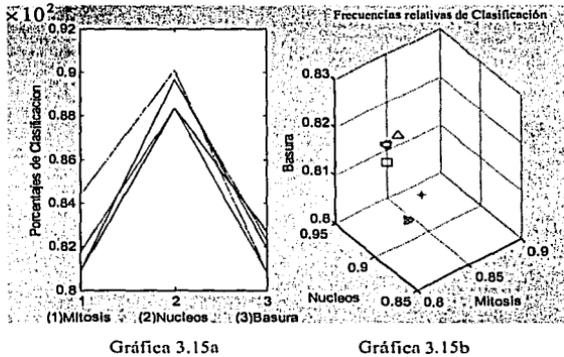


Figura 3.15 Conjunto final de CNAs no dominados de la familia uno.

En la gráfica 3.15a de la figura 3.15 podemos ver que las frecuencias relativas (porcentajes) que alcanzaron los CNAs no dominados de la familia uno están aproximadamente dentro de los intervalos:  $[0.81, 0.85]$  para patrones de mitosis,  $[0.88, 0.9]$ , para patrones clasificados como núcleos y  $[0.805, 0.83]$  para patrones tipificados como basura. De los intervalos antes mencionados podemos decir que los CNAs no dominados finales de la familia uno lograron encontrar porcentajes de clasificaciones correctas dentro de la región de porcentajes aceptables que se determinaron en la ecuación 2.2. Ahora veamos que tan parecidos, funcionalmente, son los CNAs no dominados finales de la familia uno. Las longitudes de los intervalos antes mencionados son: 0.04, 0.02 y 0.025, para mitosis, núcleos y basura, respectivamente. Las magnitudes de las longitudes de los intervalos antes mencionados, nos ilustran que las variaciones de las frecuencias relativas de aciertos de clasificación de los CNAs no dominados finales de la familia uno son muy cercanas, lo cual hace pensar que

las alternancias en los porcentajes de aciertos de clasificaciones también son muy estrechas. Ahora en la gráfica 3.15b en la figura 3.15 es más sencillo distinguir cual es la cardinalidad del conjunto final de CNAs no dominados de la familia uno. En la gráfica 3.15b podemos notar cinco puntos, en donde cada punto esta representado por un símbolo diferente. Pero, si nos fijamos con detenimiento, podemos notar que dos puntos diferentes en la gráfica 3.15b tienen símbolos que se enciman, lo cual como vimos anteriormente significa que se presentó la replicación de algunos CNAs no dominado finales, y realmente la cardinalidad del conjunto final de CNAs no dominados de la familia uno es igual a siete CNAs no dominados. Entonces en conjunto podemos decir que según nuestra propuesta de evolución de CNAs alcanzó una efectividad del 14%, ya que siete de 50 CNAs en la familia uno, lograron cumplir simultáneamente todos los objetivos determinados por la ecuación 2.2. Seleccionemos de forma aleatoria un CNA del conjunto final de CNAs no dominado de la familia uno. Veamos la figura 3.16.

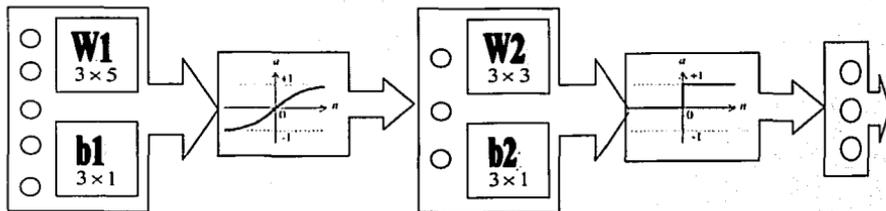


Figura 3.16 Configuración de la arquitectura del CNA no dominado seleccionado de la familia uno.

El genotipo del CNA de la figura 3.16 tiene la configuración siguiente: en la capa intermedia una función de activación tangente hiperbólica, tres neuronas en la capa intermedia y en la capa de salida una función de activación escalón positiva, además cinco neuronas en la capa de entrada y tres neuronas en la capa de salida. Pero nuevamente, el genotipo de un CNA no nos dice mucho respecto de la familia de CNAs de donde fue seleccionado. Como ya vimos en la sección 3.2.2, el fenotipo de un CNA nos dice más respecto al comportamiento de toda una familia de CNA, sobre todo en el caso del conjunto final de CNAs no dominados mostrado en la figura 3.15. Los intervalos de los porcentajes

de aciertos de clasificación que alcanzaron los CNAs no dominados de la familia uno poseen longitudes cuyas magnitudes son muy pequeñas, lo cual indica que fenotípicamente son equivalentes.

Distribución de los patrones		Resultados del Clasificador				
Clases	# de patrones	Mitosis	Núcleos	Basura	# de Errores	Frecuencia relativa de Aciertos
Mitosis	115	94	1	20	21	0.8174
Núcleos	232	1	205	26	27	0.8836
Basura	271	47	5	219	52	0.8081
Total	618	142	211	265	100	0.8382
Frecuencia relativa de Falsos Positivos		0.3380	0.0284	0.1736		
Suma del error cuadrático: 200.000000						
Error cuadrático promedio: 0.323625						

Figura 3.17 Tabla 3.17 con los valores de la matriz de confusión resultante de la etapa de validación del CNA no dominado de la figura 3.16

En la figura 3.17 podemos ver cual es el fenotipo del CNA de la figura 3.16. Es importante, reiteradamente, resaltar que el CNA de la figura 3.16 fue entrenado con una muestra no homogénea con las proporciones que se describieron al inicio de esta sección, pero su validación se llevó a cabo con la muestra de validación, la cual no tiene ningún patrón de la muestra de entrenamiento, es decir, la muestra de entrenamiento y la muestra de validación son conjuntos disjuntos. Así en la prueba de validación, a los CNAs no dominados finales de cada familia, se les presentan patrones que nunca antes les han sido presentados.

En la matriz de confusión contenida en la tabla de la figura 3.17 podemos ver que el CNA de la figura 3.16, en su etapa de validación, clasificó:

- 142 patrones como mitosis, pero 48 están mal clasificados, ya que uno verdaderamente es patrón de núcleos y 47 acertadamente son patrones tipificados como basura.
- 211 patrones como núcleos, pero seis están equivocadamente clasificados, ya que uno verdaderamente es patrón de mitosis y cinco acertadamente son patrones asociados a basura.

TESIS CON  
MALLA DE ORIGEN

- 265 patrones como basura, pero 46 están erróneamente clasificados, ya que veinte verdaderamente son patrones de mitosis y 26 acertadamente son patrones clasificados como núcleos.

Tales errores de clasificaciones dieron por resultado frecuencias relativas de falsos positivos cuya suma es igual a 0.54. Además, el valor de la varianza de las frecuencias relativas de aciertos de clasificación obtenidos por el CNA de la figura 3.16 es igual a  $1.7 \times 10^{-3}$ . Ambos valores junto con las frecuencias relativas de falsos positivos nos indican tres cosas: que el CNA de la figura 3.16 es no sesgado, proporcionalmente comete pocos errores de clasificación y que alcanza buenos porcentajes aciertos de clasificación.

Veamos ahora cual es el comportamiento del conjunto final de CNAs no dominados de la familia dos. Véase figura 3.18.

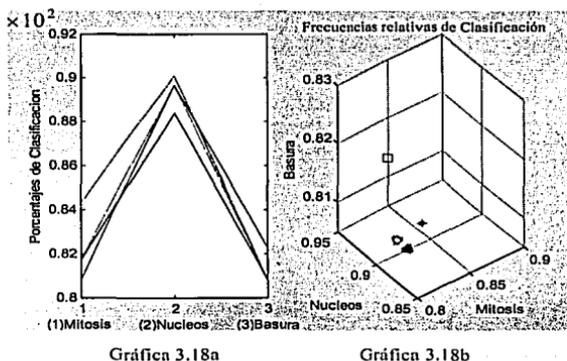


Figura 3.18 Conjunto final de CNAs no dominados de la familia dos.

De ambas gráficas de la figura 3.18, podemos obtener que la cardinalidad del conjunto final de CNAs no dominados de la familia dos es igual a seis elementos; de ellos dos pares presentan replicación. Así que la efectividad de nuestra propuesta para la evolución de CNAs en la familia dos es igual al 12%, seis de cincuenta CNAs son óptimas para la clasificación de patrones de etapas de desarrollo celular. Como ha venido siendo

costumbre, para analizar de forma particular el comportamiento de los CNAs de una familia seleccionamos aleatoriamente uno de ellos. Figura 3.19.

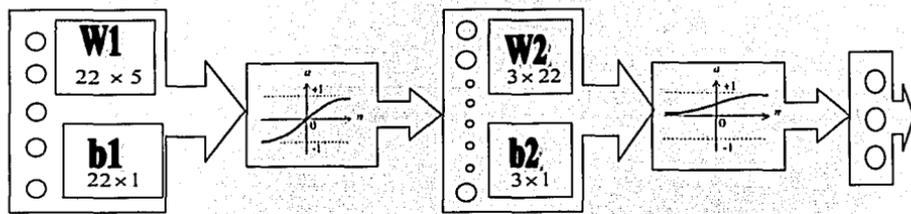


Figura 3.19 Configuración de la arquitectura de un CNA no dominado de la familia dos

El genotipo del CNA de la figura 3.19 tiene la siguiente configuración: en la capa intermedia una función de activación del tipo tangente hiperbólica, 22 neuronas en la capa intermedia y en la capa de salida una función de activación del tipo sigmoide logarítmica. Es de resaltar la magnitud del número de neuronas en la capa intermedia, veintidós, es de las mayores que se encontraron, en todos los 260 ensayos exitosos del total de 300 pruebas que se realizaron de nuestra propuesta de diseño y entrenamiento evolutivo y multiobjetivo. Veamos cual es el fenotipo del CNA de la figura 3.19.

Distribución de los patrones		Resultados del Clasificador				
Clases	# de patrones	Mitosis	Núcleos	Basura	# de Errores	Frecuencia relativa de Aciertos
Mitosis	115	93	1	21	22	<b>0.8087</b>
Núcleos	232	1	208	23	24	<b>0.8966</b>
Basura	271	43	5	223	48	<b>0.8229</b>
Total	618	137	214	267	94	0.8479
Frecuencia relativa de Falsos Positivos		0.3212	0.0280	0.1648		
Suma del error cuadrático: 283.296100						
Error cuadrático promedio: 0.458408						

Figura 3.20 Tabla 3.20 con los valores de la matriz de confusión resultante de la etapa de validación del CNA no dominado de la figura 3.19

TESIS CON  
FALLA DE ORIGEN

En la figura 3.20 se muestra la tabla que contiene la matriz de confusión de la prueba de validación del CNA de la figura 3.19. De los valores contenidos en ella podemos observar que el CNA no dominado elegido de la familia dos obtuvo los siguientes resultados:

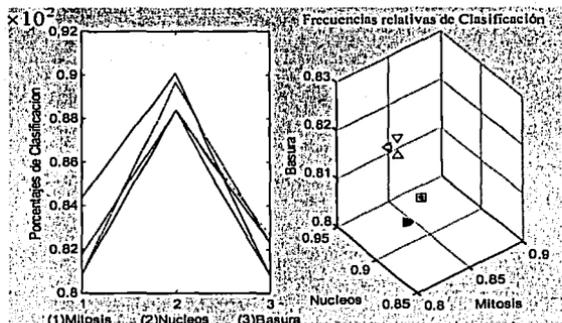
- 137 patrones los clasificó como mitosis, pero acertadamente uno es patrón de la clase núcleos y 43 son patrones de la clase basura.
- 214 patrones los clasificó como núcleos, pero correctamente uno es de la clase mitosis y cinco son patrones de la clase basura.
- 267 patrones los clasificó como basura, pero verdaderamente 21 son patrones de mitosis y 23 son patrones de la clase núcleos.

La suma de las frecuencias relativas de falsos positivos del CNA de la figura 3.19, en su etapa de validación, da por resultado el valor 0.5140 y la varianza de sus frecuencias relativas de aciertos de clasificación tiene por resultado el valor  $2.2 \times 10^{-3}$ . La magnitud de la suma de las frecuencias relativas de falsos positivos y la magnitud de la varianza de las frecuencias relativas de aciertos de clasificación nos indica, respectivamente, que son pocos los errores de clasificación y que es poco el sesgo del CNA de la figura 3.19. Además, las frecuencias relativas de aciertos de clasificación, que se muestran en la última columna de la tabla en la figura 3.20, nos indican que el CNA de la figura 3.19 alcanzó porcentajes de aciertos de clasificación dentro de la región definida por la ecuación 2.2 como región de porcentajes aceptables.

Finalmente veamos en la figura 3.21 cual es el conjunto final de CNAs no dominados de la familia tres.

En el par de gráficas de la figura 3.21 podemos notar que todos los CNAs no dominados finales alcanzaron porcentajes de clasificaciones correctas dentro de la región de porcentajes aceptables definida por la ecuación 2.2. Además podemos obtener de ambas gráficas en la figura 3.21, que la cardinalidad del conjunto de CNAs no dominados finales de la familia tres es igual a siete elementos. Al igual que para las familias uno y dos, y también cuando evolucionaron CNAs con una muestra de entrenamiento de proporciones

homogéneas; en la familia tres de CNAs evolucionados con una muestra de entrenamiento de proporciones heterogéneas, se observa el fenómeno de replicación de CNAs no dominados. Tal efecto se nota más en la gráfica 3.21b, en donde un par de puntos en el espacio tridimensional resulta ser en verdad un cuarteto de puntos.



Gráfica 3.21a

Gráfica 3.21b

Figura 3.21 Conjunto final de CNAs no dominados de la familia tres.

Seleccionemos aleatoriamente un CNA no dominado de la familia tres. Este se muestra en la figura 3.22.

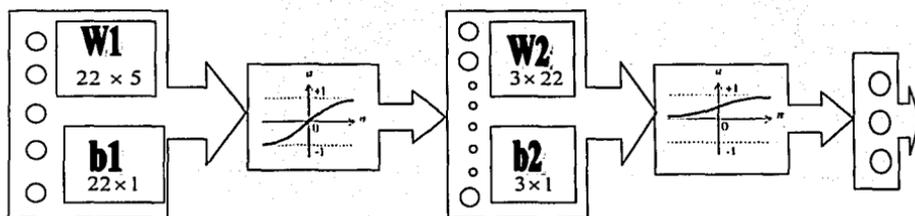


Figura 3.22 Configuración de la arquitectura de un CNA no dominado de la familia tres

Obsérvese el genotipo del CNA de la figura 3.22 y compare con el genotipo del CNA de la figura 3.19. El resultado de la comparación es obvio, aparentemente son idénticos. Esto también se presentó en varios de los ensayos que se hicieron de nuestra propuesta de evolución de CNAs. Esto se explica por el mecanismo de migración que incluimos en la modificación al algoritmo genético simple que propusimos en el capítulo dos. La migración que estamos proponiendo es sin restricciones, en el sentido de que los elementos mejor adaptados de una familia pueden migrar a cualquiera de las otras familias y sustituir a los elementos menos adaptados. De forma simplificada, el mecanismo de migración representa para nosotros un elemento de perturbación que tiene por objetivo traer variabilidad en el desarrollo de las tres familias de CNAs y atenuar lo efectos de la deriva genética y la endogamia. Veamos cual es el fenotipo del CNA de la figura 3.22, veamos su matriz de confusión contenida en la tabla de la figura 3.23.

Distribución de los patrones		Resultados del Clasificador				
Clases	# de patrones	Mitosis	Núcleos	Basura	# de Errores	Frecuencia relativa de Aciertos
Mitosis	115	97	1	17	18	<b>0.8435</b>
Núcleos	232	1	210	21	22	<b>0.9052</b>
Basura	271	42	10	219	52	<b>0.8081</b>
Total	618	140	221	257	93	0.8495
Frecuencia relativa de Falsos Positivos		<b>0.3071</b>	<b>0.0498</b>	<b>0.1479</b>		
Suma del error cuadrático: 262.323625						
Error cuadrático promedio: 0.4245						

Figura 3.23 Tabla 3.23 con los valores de la matriz de confusión resultante de la etapa de validación del CNA no dominado de la figura 3.22

Comparando las tablas en las figuras 3.20 y 3.23 es obvio que los valores de los porcentajes de aciertos de clasificación son distintos y en general son diferentes ambas matrices de confusión. La suma de las frecuencias relativas de falsos positivos del CNA de la figura 3.22 es igual a 0.5048 y la varianza de las frecuencias relativas de aciertos de clasificación es igual a  $2.4 \times 10^{-3}$ . Estos valores son diferentes a los obtenidos por el CNA de la figura 3.19. ¿Por qué fenotípicamente son diferentes los CNAs de las figuras 3.19 y 3.22, si genotípicamente son idénticos? En las figuras 3.19 y 3.22 y en general en todas las gráficas en donde mostramos el genotipo de un CNA, solo hemos mostrado la parte más superficial,

por así decirlo, de la configuración genotípica. El genotipo de un CNA no sólo lo componen las funciones de activación y el número de neuronas de todas y cada una de las capas. En todas las ilustraciones de los genotipos de CNAs que hemos analizado también se han incluido las matrices de pesos y los términos de tendencia, que en la mayoría de los CNAs analizados no han marcado la diferencia entre un CNA y otro. Pero en el caso de los CNAs de las figuras 3.19 y 3.22 la variación de algunos valores en las matrices de pesos y/o en los vectores de tendencia explican por que las matrices de confusión de los CNAs de las figuras 3.19 y 3.22 son diferentes. A este respecto no existe relación alguna que nos indique de forma determinística qué neuronas y por consiguiente qué pesos o términos de tendencia significan la diferencia.

Es fácil de predecir correctamente, que en varios de los ensayos de nuestra metodología evolutiva y multiobjetivo de diseño y ajuste o entrenamiento de CNAs, se dio el caso de que un mismo CNA no dominado estuviera en más de una de las familias y que como en el caso anterior, los CNAs aparentemente con el mismo genotipo, tuvieran distinto fenotipo; o que tuvieran el mismo genotipo y fenotipo aún perteneciendo a familias diferentes.

Número de Figura del CNA de referencia	Cantidad Total de errores de clasificación	Suma del Error Cuadrático	Error Cuadrático Promedio	Varianza de las frecuencias relativas de Aciertos de clasificación	Suma de las frecuencias relativas de Falsos positivos
Figura 3.16	100	200.0	0.323625	$1.7 \times 10^{-3}$	0.54
Figura 3.19	94	283.296100	0.458408	$2.2 \times 10^{-3}$	0.514
Figura 3.22	93	262.323625	0.4245	$2.4 \times 10^{-3}$	0.5048

Figura 3.24 Tabla 3.24 Resumen de los valores obtenidos, en la etapa de validación por los CNAs analizados anteriormente evolucionados con una muestra de entrenamiento de proporciones heterogéneas.

Finalmente comparemos los CNAs no dominados que hemos analizado y fueron evolucionados con una muestra de entrenamiento de proporciones heterogéneas. Para esto veamos la tabla en la figura 3.24.

Nuevamente el error cuadrático promedio no es una buena métrica para medir el desempeño de los CNAs, porque, ya que como podemos ver, según la tabla 3.24, el CNA con el menor error cuadrático promedio no fue quién obtuvo una cantidad menor de errores

de clasificación. Incluso en un caso de la muestra de proporciones heterogéneas, este hecho es más evidente que cuando se evolucionó con una muestra de entrenamiento de proporciones homogéneas. Así nuevamente se comprueba que la evaluación del desempeño de un CNA no permite la existencia de métrica absoluta y unidimensional, es más conveniente establecer composiciones simultáneas de métricas relativas para poder jerarquizar el desempeño en conjuntos de CNAs.

### 3.3 COMPARACIÓN CON OTROS MÉTODOS

De alguna manera, nuestra propuesta de modificación al algoritmo genético simple, se podría percibir como una variante al método no basado en Pareto propuesto por David Schaffer en 1985 llamado VEGA por las siglas de Vector Evaluated Genetic Algorithm. VEGA es una extensión directa del Algoritmo Genético Simple. El procedimiento es muy sencillo, para una función con  $M$  objetivos a optimizar, se particiona aleatoriamente la población total de individuos en  $M$  subpoblaciones de igual tamaño. Los elementos de cada subpoblación tendrán una medida de desempeño en base a una función objetivo. Entonces cada subpoblación tendrá una distinta función de desempeño (fitness). De tal forma que cada una de las  $M$  funciones objetivo es empleada para evaluar a algunos elementos en la población. El alcance del operador de selección se limita únicamente a que opere en cada subpoblación. Posteriormente se mezclan aleatoriamente todas las subpoblaciones para obtener nuevamente una sola población total. Sobre la población total se desarrollan los operadores de cruzamiento y mutación. Nuevamente se particiona aleatoriamente la población total de individuos en  $M$  subpoblaciones de igual tamaño y se repite el proceso.

Con el objetivo de evaluar la supuesta similitud entre VEGA y nuestra modificación al algoritmo genético simple, programamos VEGA y lo adaptamos para el diseño y ajuste de CNAs. Algunos resultados se muestran en las siguientes figuras.

En la figura 3.25 podemos ver la matriz de confusión de un CNA entrenado con VEGA. El genotipo de este CNA tiene en la capa intermedia ocho neuronas y las funciones de activación escalón simétrica y la función lineal saturada en las capas intermedia y de salida, respectivamente. El CNA cuya matriz de confusión esta contenida en la tabla de la figura 3.25, fue evolucionado con una muestra de entrenamiento de proporciones homogéneas.

Como podemos ver, en la tabla de la figura 3.25, no todos los porcentajes de aciertos de clasificación tienen un valor superior al 80%, a pesar de que el número total de errores es igual a 97 incluso menor que el total de errores de clasificación que cometieron los CNAs de las figuras 3.1 y 3.16 es igual al CNA de la figura 3.4. De la tabla 3.25 podemos obtener que la suma de las frecuencias relativas de falsos positivos es igual a 0.4481, cuyo valor es menor al mismo factor en todos los CNAs evolucionados con una muestra de entrenamiento de proporciones heterogéneas, inclusive menor que la suma de las frecuencias relativas de falsos positivos del CNA de la figura 3.1.

Distribución de los patrones		Resultados del Clasificador				
Clases	# de patrones	Mitosis	Núcleos	Basura	# de Errores	Frecuencia relativa de Aciertos
Mitosis	127	98	1	28	29	<b>0.7717</b>
Núcleos	276	0	256	20	20	<b>0.9275</b>
Basura	332	30	18	284	48	<b>0.8554</b>
Total	735	128	275	332	97	0.8680
Frecuencia relativa de Falsos Positivos		<b>0.2344</b>	<b>0.0691</b>	<b>0.1446</b>		
Suma del error cuadrático: 166.632362						
Error cuadrático promedio: 0.226711						

Figura 3.25 Tabla 3.25 con los valores de la matriz de confusión resultante de la etapa de validación de un CNA entrenado con la técnica de VEGA (Ejemplo uno).

Distribución de los patrones		Resultados del Clasificador				
Clases	# de patrones	Mitosis	Núcleos	Basura	# de Errores	Frecuencia relativa de Aciertos
Mitosis	111	85	3	23	26	<b>0.7658</b>
Núcleos	196	1	190	5	6	<b>0.9694</b>
Basura	237	20	44	173	64	<b>0.7300</b>
Total	544	106	237	201	96	<b>0.1765</b>
Frecuencia relativa de Falsos Positivos		<b>0.1981</b>	<b>0.1983</b>	<b>0.1393</b>		
Suma del error cuadrático: 229.165276						
Error cuadrático promedio: 0.421260						

Figura 3.26 Tabla 3.26 con los valores de la matriz de confusión resultante de la etapa de validación de un CNA entrenado con la técnica de VEGA (ejemplo dos)

También se probó evolucionar CNAs mediante VEGA con una muestra de entrenamiento de proporciones heterogéneas. La matriz de confusión del CNA resultado se muestra en la

TESIS CON  
FALLA DE ORIGEN

tabla de la figura 3.26. En ella pudimos observar que existe sesgo al clasificar mejor núcleos que mitosis y basura. No obstante que la suma de las frecuencias relativas de falsos positivos es igual a 0.5357, este valor es menor a la suma de las frecuencias relativas de falso positivos obtenida por el CNA de la figura 3.16.

Este tipo de matrices de confusión que denotan sesgo, son los ejemplos típicos que reportó la serie de ensayos realizados con VEGA. Esto se explica por la razón que la técnica de Schaffer tiene la tendencia a eliminar individuos con desempeño promedio en todas las dimensiones. Esto no sucede con el ordenamiento de Fonseca, el cual contrariamente tiende a favorecer en alguna medida la sobrevivencia de CNAs con porcentajes de clasificaciones promedio, los cuales tienen una importante potencialidad de convertirse en CNAs óptimos en la medida que se den las subsecuentes evoluciones del plan evolutivo.

Además de VEGA probamos con otras técnicas: El ordenamiento lexicográfico y el uso de una función de agregación. El ordenamiento lexicográfico nos dio en mayor medida CNAs sesgados, ya que se observó que una vez que un CNA ha optimizado el porcentaje de aciertos de clasificación de una clase, es muy difícil que después logre remontar y optimizar el resto de los porcentajes de clasificación de las otras clases sin que pierda lo ganado anteriormente. Esto es debido a que el ajuste de peso se hace de forma aleatoria y no se trata de epochs como en los métodos de entrenamiento neuronales clásicos.

En función de agregación de inmediato nos causó problemas poder situar la población inicial de CNAs dentro de una región de valores factibles para los distintos objetivos. Además, dada la disparidad de las escalas de los distintos objetivos que establece la función de la ecuación 2.2, nos causó problemas lograr calibrar el vector de los coeficientes de ponderación debido a que no sabemos qué objetivo tiene mayor importancia: la varianza de las frecuencias relativas de aciertos de clasificación, la suma de las frecuencias relativas de falsos positivos o alguno de los porcentajes de aciertos de clasificación.

En cuanto a las metodologías clásicas es perfectamente conocido que permiten diseñar CNAs robustos que pueden resolver problemas de clasificación difíciles y complejos de forma confiable. Sin embargo, para poder obtener ese tipo de CNAs se tiene que seguir un procedimiento largo, en el cual no son claras las reglas que garantizan obtener el mejor CNA posible. Es necesario tomar decisiones en torno al diseño del CNA en base a la experiencia o de manera empírica. Estos factores dificultan establecer un proceso

sistemático y general de diseño y obligan a seguir diferentes procedimientos de diseño dependientes de cada problema específico. En consecuencia, el diseño de un CNA se convierte en un proceso largo y tedioso. El proceso clásico de diseño y ajuste de un CNA puede tomar semanas o todavía meses de trabajo para conseguir un buen CNA. Tenemos que recordar que estamos ante un problema de combinatoria, donde se podría estar probando todas las posibles combinaciones de configuraciones de arquitecturas de CNAs y métodos de entrenamiento para determinar cual combinación produce los mejores CNAs. En el siguiente capítulo discutiremos las conclusiones a las que nos llevaron la serie de resultados analizados en este capítulo, las conclusiones particulares de algunos elementos de la metodología y las conclusiones de la eficacia de la metodología propuesta. También se analizan las conclusiones sobre las características propias del problema de clasificaciones y finalmente las conclusiones puntuales del estadístico desarrollado para la prueba de la metodología.

TESIS CON  
FALLA DE ORIGEN

TESIS CON  
FALLA DE ORIGEN

## CAPÍTULO CUATRO

### CONCLUSIONES

#### 4.1 DISCUSIÓN

El número de posibles parámetros para el desarrollo y entrenamiento de un clasificador artificial plantea un espacio de búsqueda cuya explosión combinatoria es excesiva. Con esta premisa, los resultados anteriores demuestran que una población de soluciones potenciales, específicamente, una población de tres familias de CNAs tiene más factibilidad de encontrar buenos CNAs que una metodología basada en una serie de pruebas de calibración con un CNA individual. Esta medida de factibilidad se traduce en que a partir del desarrollo y entrenamiento de la población de familias de CNA, es más probable encontrar algún CNA o un conjunto de CNA con una configuración de diseño óptimo cuyo comportamiento fenotípico sea sin sesgo, sin bajo ni sobre ajuste, con buenos porcentajes de clasificación para la muestra de validación. Pero analicemos con más detalle las conclusiones

#### 4.2 MUESTREO DEL CONJUNTO DE ENTRENAMIENTO Y DEL CONJUNTO DE VALIDACIÓN

Como se planteó en el capítulo uno y se especificó en el capítulo dos, la forma en que se selecciona una muestra para entrenar los CNA es muy importante. El preprocesamiento que se haga a los datos es un factor a considerar que afecta el diseño y entrenamiento de los CNAs. En la metodología hemos propuesto cómo realizar la selección de una muestra de entrenamiento que sea representativa de la muestra experimental. Esto lo logramos definiendo intervalos típicos muestrales, para las cinco características ( Forma, Área, Nivel de gris, Dep. Hoyos y DENG ) de las tres clases de patrones de entrenamiento. Pero no sólo es importante cómo se selecciona, sino también cuánto se selecciona. Las proporciones son importantes; sabemos que el tamaño de las muestras de entrenamiento y de validación es un factor a tomarse en cuenta en los métodos clásicos de entrenamiento neuronal. Una muestra de entrenamiento que tenga mayor representatividad proporcional de una clase determinada, como es nuestro caso, puede orillar a entrenar los CNAs con tendencia a clasificar mejor esa determinada clase. En el problema particular que estamos resolviendo, las muestras experimentales reales por su naturaleza misma tienen un sesgo implícito en las

proporciones, favorable a los patrones basura en relación a mitosis y núcleos. Los resultados demuestran que si el diseño y ajuste de CNAs es planteado en términos de las técnicas evolutivas y multiobjetivo, no es necesario evitar las características propias del problema y por lo tanto no es necesario tratar de balancear las proporciones, ya que nuestra metodología incluye la selección automática de muestras representativas sin importar las proporciones de sus distribuciones.

Así, desde la selección adecuada de las muestras de entrenamiento es posible evitar en cierta medida el sesgo, con objeto de poder diseñar y ajustar, evolutivamente, CNAs con porcentajes de clasificación aceptables tanto para la muestra de entrenamiento como para la de validación. Esto se observó en los CNAs no dominados del capítulo tres. El porcentaje de abatimiento de los porcentajes de aciertos de clasificación no fue mayor al 10% en los ensayos exitosos, indicando con ésto que los CNAs no dominados obtenidos poseen una buena medida de generalización, lo cual es muy importante cuando los CNAs obtenidos clasifiquen nuevas muestras experimentales de datos.

### 4.3 FUNCIONES DE ACTIVACIÓN

En todo el estadístico desarrollado presentado en el capítulo tres del conjunto de funciones de activación mostrado en la figura 2.2, la variedad de configuraciones en el diseño de los CNAs incluyeron la mayor parte del conjunto de las funciones definidas en una amplia variedad de combinaciones. Pero encontramos una singularidad, de todas las funciones de activación definidas, la función base radial que tiene la ecuación 4.1 y toma forma gráfica en la figura. 4.1, no apareció en ninguna configuración de los CNAs no dominados finales en ninguno de los ensayos exitosos. Esto nos hace concluir que el problema particular de clasificación de patrones de estados de desarrollo celular ( mitosis, núcleos y basura ) requiere de una configuración óptima de matrices de pesos, términos de tendencia y una combinación de funciones de activación que incluya solo funciones del tipo monótonamente crecientes.

$$a(x) = e^{-x^2}$$

Ecu. 4.1.

De los resultados del estadístico de prueba de nuestra metodología, podemos afirmar que la combinación de funciones de activación no es un determinante para el diseño de CNAs de patrones de estado celulares, ya que encontramos una amplia variedad de CNAs con buenos porcentajes de clasificación que genotípicamente son muy diferentes, pero fenotípicamente son muy similares. Esto es, estructuralmente tienen una configuración de matrices de pesos, vectores de tendencia y combinación de funciones de activación, completamente distintas, pero que funcionalmente son muy semejantes. Los porcentajes de clasificación son muy cercanos, y en todos el 86% de los ensayos caen dentro de la región aceptable de porcentajes de aciertos de clasificación.

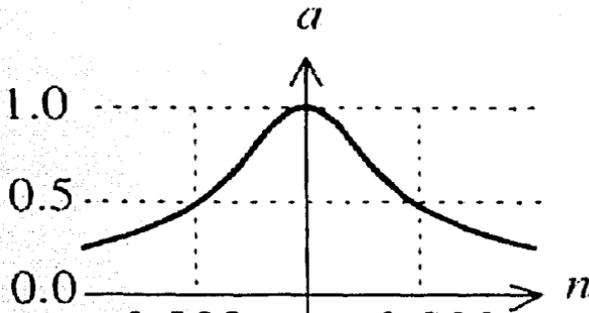


Figura 4.1 Función de activación Base Radial

#### 4.4 SOBRE Y BAJO AJUSTE.

Es obvio que el número de neuronas determina el número de pesos en cada CNA. Cada peso es un parámetro que incrementa o decrementa la capacidad de clasificación de los CNAs. Entonces, evitar los efectos del sobre y bajo ajuste, significa calibrar el número de neuronas en el CNA. En las familias de CNAs, todas las configuraciones tienen cinco neuronas en la capa de entrada y tres neuronas en la capa de salida. Esto es, por demanda particular del problema de clasificación de patrones de etapas de desarrollo celular, ya que tenemos tres clases de patrones (mitosis núcleos y basura) y a su vez cada patrón tiene cinco características forma (área, nivel de gris, Dep. hoyos y D.E.N.G.). Los operadores

de cruzamiento y mutación se encargan de realizar los procesos de poda y crecimiento neural de manera un tanto estocástica, pero siempre obedeciendo las restricciones y direcciones que le marca la función multiobjetivo definida en la ecuación 2.2. En particular es concluyente notar que, en la totalidad de los ensayos exitosos la dimensión de la capa intermedia estuvo acotada entre tres y veintidós neuronas, no obstante que el rango de variabilidad que se le dio a este parámetro de diseño fue de 1 a 32 neuronas.

Obviamente que no podemos decir que los términos en que se diseñaron los operadores de cruzamiento y mutación tienen todo el crédito de evitar el sobre y bajo ajuste. En gran medida fue una labor de equipo; esto quiere decir, que los operadores de cruzamiento y mutación en conjunción con el uso de tres familias de CNAs independientes entre si, y la propuesta de función multiobjetivo, coordinada con la técnica de Fonseca en combinación con el procedimiento de migración sin restricciones, son elementos suficientes para, primero dar variedad a los CNA dominantes resultantes. Esto se refleja en que el efecto de la deriva genética, la endogamia y la replicación se atenúan lo suficiente. Segundo, se obtienen CNAs dominantes balanceados, es decir, sin sesgo. Tercero, se obtienen conjuntos de CNAs con capacidades de generalidad y similares fenotípicamente, lo cual significa que estamos encontrando conjuntos robustos de CNAs.

#### 4.5 FUNCIÓN OBJETIVO

A fin de evaluar correctamente el desempeño de CNAs, es importante invertir el tiempo suficiente en el diseño de una función objetivo adecuada. En general la función objetivo determina en gran medida cual debe ser y la forma en que debe de plantearse la solución de un problema. La función objetivo es el factor directriz, nos indica la calidad de la solución que estamos proponiendo. A este respecto una de las mejores opciones a considerar es tomar en cuenta las características del problema. En particular para los CNAs es importante considerar tres factores:

- Provocar el mayor número de aciertos de clasificaciones correctas.
- Esquivar en lo posible el sesgo hacia una o unas determinadas clases.

- Evitar cometer errores de clasificación ( falsos positivos ).

Como podemos ver los tres factores anteriores no son independientes entre sí, además que sus relaciones no son directas. Así la función multiobjetivo que proponemos explora en ese camino, ya que de forma simultánea:

- Se considera la maximización de los porcentajes de aciertos de clasificación, pero en un rango aceptable determinado por un intervalo que procura en sus cotas inferior y superior evitar el sobre y bajo ajuste, respectivamente.
- Se pondera la minimización del número de falsos positivos que un CNA comete, en general durante su desempeño, en base a reducir sistemáticamente el valor de la sumatoria de los porcentajes de falsos positivos.
- Y finalmente se estima esquivar el fenómeno del sesgo mediante la minimización de la varianza sobre los porcentajes de los aciertos de clasificación.

Así, este sentido, la función objetivo debe tomar en cuenta métricas relativas y simultáneas, con relaciones complejas que consideren las características de las distribuciones de las muestras de patrones de desarrollo celular y las características propias de las CNAs.

#### 4.6 METODOLOGÍA

De estas conclusiones y los resultados anteriores podemos afirmar que, la funcionalidad de la metodología que implementamos es suficientemente eficiente para el problema particular de clasificación de patrones de estados celulares. Como vimos en cada familia de CNAs, el porcentaje de CNAs no dominados finales está alrededor del 10%, lo cual ya es una ventaja frente a la forma tradicional de diseñar y entrenar CNAs.

El resultado anterior debe tomarse considerando que la metodología incluye automáticamente la fase final del preprocesamiento a los datos experimentales. Esto es, realizamos nuestra propia propuesta de muestreo con el fin de obtener representatividad en la muestra de entrenamiento, lo cual significa para el éxito final de los CNAs en la etapa de

validación, que la probabilidad de encontrar buenos porcentajes de aciertos de clasificación se incrementa.

La metodología que proponemos también resuelve el problema de deriva genética y replicación fenotípica. El diseño de los operadores de cruzamiento y mutación que proponemos logra diversidad en las etapas iniciales del desarrollo de las familias de CNA y, el proceso de migración aleatoria no restringida logra dar mayor dispersión genotípica y fenotípica al realizar el intercambio de información cromosómica entre las familias de CNAs.

Como se dijo en el capítulo dos, un factor crucial es el proceso de definición y selección de los mejores CNAs. La determinación de una función multiobjetivo que plantea una heurística de selección, muy propia del problema de clasificación en general, es clave para poder obtener CNAs aceptables en sus porcentajes de clasificación. La inclusión de la técnica de Fonseca también es vital, ya que nos enriquece la técnica de Pareto, en donde el ordenamiento únicamente está determinado por la no dominancia. Y nos evita la problemática de VEGA y otras técnicas. Nuestro enfoque trae consigo que CNAs mediatizados tengan la posibilidad de ser CNAs no dominados y enriquecer el desarrollo evolutivo individual de las tres familias. Como vimos, la técnica de Fonseca trae mayor diferenciación selectiva. Esto trae consigo diversidad en las familias de CNAs y asegura, en cierta medida, dispersión poblacional.

En todos los casos los objetivos se cumplen, se toman en cuenta las características y restricciones relativas al problema específico a resolver. La metodología incluye la capacidad de seleccionar automáticamente una muestra representativa. Es capaz de diseñar y ajustar CNAs con una amplia variedad de configuraciones, pero con funcionalidades equivalentes. Esto proporciona una medida de robustez a nuestra metodología, podemos dar diferentes soluciones para el problema de clasificación de patrones de estado de desarrollo celular.

El trabajo expuesto de forma específica tiene tendencia natural a ser una continuación de un trabajo más amplio, al unir dos grandes áreas de investigación: las redes neuronales y las técnicas evolutivas. Las vertientes de trabajo futuro son amplias, puede ser tomado como una herramienta o laboratorio virtual para realizar estudios más específicos en ambas áreas. En las redes neuronales, probar con otra variedad de problemas, además de continuar el

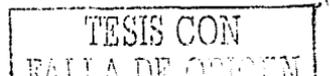
estudio de los clasificadores. Desde el punto de vista de las técnicas evolutivas, los parámetros evolutivos son siempre susceptibles de calibración para optimizar los resultados. La combinatoria en este sentido también es fuerte y con el proceso de migración artificial ampliamos las perspectivas de estudio; esto nos da oportunidad de comenzar a planear un conjunto de pruebas para avanzar en este sentido.

En resumen, con la programación elaborada para sustentar las afirmaciones de este trabajo, tenemos una herramienta que puede realizar todo el trabajo árduo ( evaluar, seleccionar y adaptar los CNAs hasta un punto donde obtenemos un conjunto de CNAs óptimos, en donde el total de tiempo necesario para tal fin, es del orden de un par de horas como máximo, dependiendo del tamaño de la muestra de datos), únicamente tenemos que diseñar la estructura cromosómica necesaria y calibrar el algoritmo evolutivo.

Finalmente, dos aspectos importantes deben ser remarcados al respecto de los CNAs obtenidos mediante nuestra propuesta. El primero es que, en base a los conjuntos de CNAs óptimos encontrados, no existe una topología óptima específica para el CNA óptimo. Esto sugiere que podemos elegir una configuración arbitraria de arquitectura de CNA y con un entrenamiento adecuado hallar un buen CNA. Esto abre un área interesante de investigación en el sentido de analizar y delimitar la veracidad de esta idea.

El otro aspecto importante está relacionado con el entrenamiento del CNA, debido a que con las técnicas evolutivas y multiobjetivo las ideas de entrenamiento y adaptación toman un significado diferente y se trasladan a términos como evolución y generación. Los pesos y términos de tendencia no son ajustados de acuerdo al error de clasificación. Aquí los pesos y términos de tendencia son genéticamente modificados durante el proceso evolutivo, además de la configuración de la arquitectura de los CNAs, así que al final de este proceso tenemos un CNA completo y operacional, que no necesita de entrenamiento adicional. En este caso, los conceptos de sobre y bajo entrenamiento toman otro significado debido a que no estamos entrenando en el sentido clásico. La decisión de cuando un CNA ha evolucionado lo suficiente para hacer un buen trabajo depende de las condiciones establecidas en el proceso evolutivo.

Las expectativas en general de continuidad de este trabajo son amplias. Los problemas que se pueden abordar con nuestra propuesta son aplicables al diseño automático de cualquier tipo de estructuras genéricas. La teoría de redes nos puede plantear problemáticas tan



diversas, susceptibles de ser solucionadas con las técnicas evolutivas multiobjetivo mostradas en este trabajo. Incluso, las oportunidades de aplicación de las ideas y herramientas desarrolladas en el transcurso de este trabajo ya están probándose actualmente en áreas de diseño óptimo industrial y han sido tomadas en cuenta en áreas de diseño de sistemas adaptables de diagnóstico clínico y del área de las telecomunicaciones. Desde el punto de vista teórico, este trabajo plantea inquietudes en el análisis teórico detallado de la conjunción de los efectos de los procesos de regulación artificial y la toma de decisiones multicriterio.

Los resultados de este trabajo nos impulsan a continuar estudiando las técnicas evolutivas multiobjetivo aplicadas al diseño de redes neuronales en general, en el sentido de diseñarlas y ajustarlas en una forma sencilla, fácil, rápida y sistemática, independientemente de las características del problema.

## REFERENCIAS

- [ARÁ01] Arámbula Cosío F., Vega L., Herrera Becerra A., Prieto Meléndez R., Corkidi G.  
Automatic identification of metaphase spreads and nuclei using neural networks  
Medical & Biological Engineering & Computing  
Volume 39  
Year 2001  
pp. 391-396
- [BLA00] Blanco A., Delgado M., Pegalajar M.C.  
A Real-Coded Genetic Algorithm for Training Recurrent Neural Networks  
Neural Networks, Pergamon  
Volume 14  
Año 2001  
pp 93-105
- [BÄC97] Bäck Thomas, Fogel David B., Michalewicz Zbigniew  
Handbook of Evolutionary Computation  
Institute of Physics Publishing and Oxford University Press.  
Release 1997  
USA.
- [CHI95] Chipperfield Andre, Fleming Peter, Pohlheim Hartmut, Fonseca Carlos.  
Genetic Algorithm TOOLBOX for use with MATLAB User's Guide  
Department of Automatic Control and Systems Engineering.  
University of Sheffield  
England 1995
- [CHI96] Chin-Teng Lin, Lee George C.S.  
Neural Fuzzy Systems  
A Neuro-Fuzzy Synergism to Intelligent Systems  
Prentice Hall  
NJ, USA  
1996.
- [CHE99] Chen Mei, Kanade Takeo, Pomerleau Dean, Rowley Henry A.  
Anomaly Detection Through Registration.  
Pattern Recognition  
Volume 32,  
Número 1  
January 1999.

TESIS CON  
FALLA DE ORIGEN

- [COE99] Coello Coello Carlos A.  
A Comprehensive Survey of Evolutionary-Based Multiobjective Optimization Techniques.  
Knowledge and Information Systems.  
Volume 1.  
Año 1999.  
pp 269-308
- [COR98] Corkidi G., Vega L., Márquez J., Ostrosky-Wegman P.  
Roughness feature of Metaphase chromosome spreads and nuclei for automated cell proliferation analysis  
Medical & Biological Engineering & Computing  
Volume 36  
Year 1998  
pp. 679-685
- [DAR00] Darzynkiewicz Zbigniew, Crissman Harry A., Robinson J. Paul  
Methods in Cell Biology  
Primera edición  
Academic  
San Diego  
Year 2000
- [DEB89] Deb Kalyanmoy and Goldberg David.  
An Investigation of Niche and Species Formation in Genetic Function Optimization.  
ICGA '89.  
ICGA.  
The University of Alabama.  
Tuscaloosa Alabama, USA.  
Año 1989.
- [DEB01] Deb Kalyanmoy  
Multi-Objective Optimization using Evolutionary Algorithms  
Wiley  
Sussex, England  
Año 2001
- [FON95] Fonseca Carlos.  
Multiobjective Genetic Algorithms with Application to Control Engineering Problems.  
Thesis submitted for candidate for the degree of PhD.  
The University of Sheffield.  
Department of Automatic Control and Systems Engineering.  
England.  
Año 1995.

- [FRE97] Freeman, James A., Skapura, David M.  
Redes Neuronales: algoritmos, aplicaciones y técnicas de programación  
Díaz de Santos,  
Madrid España  
1997
- [GOL89] Goldberg David E.  
Genetic Algorithms in search, Optimization and Machine Learning.  
Primera edición.  
Addison Wesley.  
USA.  
Año 1989.
- [GAR95] Garduño Ángeles Edgar  
Sistema Automático para el Cálculo del Índice Mitótico por el Procesamiento de  
Digital Imágenes MIT.  
Tesis de Ingeniería en Computación.  
Facultad de Ingeniería.  
Universidad Nacional Autónoma de México.  
México D.F.  
Mayo 1995
- [GOL99] Golubsky Wolfgang and Feuring Thomas.  
Evolving Neural Networks Structures by Means of Genetic Programming.  
EuroGP'99.  
R. Poli.  
Springer-Verlag.  
Berling Alemania  
Año 1999.  
pp. 211-220
- [GON78] Gonzalez C. Rafael, Thomason G. Michael  
Syntactic Pattern Recognition an introduction  
Addison-Wesley Publishing, Inc.  
Massachusetts, USA  
1978.
- [HAY98] Haykin Simon  
Neural Networks : A Comprehensive Foundation  
2nd Edition  
Prentice Hall.  
USA  
July 1998

TESIS CON  
FALLA DE ORIGEN

- [KUR00a] Kuri Angel  
 Training Neural Networks Using Non-standard Norms – Preliminary Results  
 MICA1 2000 Advances in Artificial Intelligence  
 Proceedings Mexican International Conference on Artificial Intelligence  
 Acapulco, México  
 April 2000  
 pp 350-364
- [KUR00b] Kuri Angel  
 Non-standard Norms in Genetically Trained Neural Networks  
 2000 IEEE Symposium on Combinations of Evolutionary Computation and  
 Neural Networks.  
 San Antonio, TX USA  
 May 2000  
 pp 43-53
- [KUR01] Kuri Angel  
 Entrenamiento de Redes Neuronales usando Algoritmos Genéticos, notas de la  
 presentación del 9 de marzo de 2001  
 Centro de Investigación en Computación  
 Instituto Politécnico Nacional  
 Seminario de Computación organizado por el Posgrado en Ciencia e Ingeniería de  
 la Computación.  
 Auditorio del IIMAS, Ciudad Universitaria, México D.F.  
 Ciclo 2001
- [JAV98] Javadzadeh, T., Jalali-Nadoshan, M.R., Shokouhi, N.  
 Correlation of the mitotic index with estrogen receptors in invasive breast cancer  
 Research Journal Pejouhandeh Vol.3 , No.1 ( Spring 1998 )  
 Editorial borrador of Shaheed Beheshti Medical University  
 Tehran - P.O.Box : 19395-3955  
 Year 1998
- [LAW94] Law Diane and Miikkulainen Risto.  
 Grounding Robotic with Genetic Neural Networks.  
 Technical Report.  
 Department of Computer Sciences.  
 The University of Texas at Austin.  
 Texas, USA.  
 Año 1994.

- [LEO98] Leondes Cornelius T.  
Implementation Techniques  
Neural Networks Systems and Applications  
Academia Press  
San Diego California, USA.  
Año 1998.
- [MAS93] Masters Timothy.  
Practical Neural Networks Recipes in C++.  
Primera edición.  
Academic Press, Inc.  
San Diego California, USA.  
Año 1993.
- [MAT92] MathWorks, Inc.  
High-Performance Numeric Computation and Visualization Software, Reference  
Guide  
MathWorks  
Natick, Mass.  
USA 1992
- [MCQ97] McQuesten Paul and Miikkulainen Risto.  
Culling and Teaching in Neuro-Evolution.  
Proc. 7<sup>th</sup> Intl. Conf. On Genetic Algorithms.  
Morgan Kaufmann.  
ICGA97.  
San Francisco, USA.  
Año 1997.
- [PAR00] Park H., Amari S.-I, Fukumizu K.  
Adaptive natural gradient learning algorithms for various stochastic models.  
Neural Networks Pergamon.  
Volume 13.  
2000.  
pp 755-764.
- [PHA99] Phan D.T., Karaboga D.  
Training Elman and Jordan networks for system identification using genetic  
algorithms  
Artificial Intelligence in Engineering.  
Volume 13.  
Año 1999.  
pp 107-117.

TESIS CON  
FALLA DE ORIGEN

- [PRI96] Prieto M. Rafael, Caballero H. Grissel, Herrera B. A. Arturo, Corkidi B. Gabriel.  
Diseño de un Clasificador Neuronal para el reconocimiento de Imágenes Biomédicas.  
Reporte Técnico.  
RTSTN-9614.  
Universidad Nacional Autónoma de México.  
Centro de Instrumentos.  
Año 1996.
- [PRI97] Prieto M. Rafael, Herrera B. A. Arturo, Corkidi B. Gabriel, Perez J. L., Quintana S.  
Design of a Neural Classifier for an Automatic Counting Cells System.  
Sian ka'an 97, International Conference.  
The Second Mexico-US Joint International Workshop on Neural Networks and Control.  
pp. 156-161.  
México  
Año 1997
- [PRI00a] Prieto M. Rafael, Herrera B. A. Arturo.  
Método de Reducción de Parámetros para la construcción de Clasificadores Neuronales  
Informe Técnico.  
Ingeniería Neuronal-9907.  
Universidad Nacional Autónoma de México.  
Centro de Instrumentos.  
Año 2000.
- [PRI00b] Prieto M. Rafael, Herrera B. A. Arturo.  
Cálculo del índice mitótico con la ayuda de un sistema Neuronal usando pocos parámetros  
Informe Técnico.  
Ingeniería Neuronal-9909.  
Universidad Nacional Autónoma de México.  
Centro de Instrumentos.  
Año 2000.
- [PRI03] Prieto M. Rafael, Rosendo R. J. Alejandro, Herrera B. A. Arturo, Padrón G. Alejandro.  
Comparison Between Two Different Methodologies to Design Neural Classifiers  
Proceedings of The 7<sup>th</sup> World Multiconference on Systemics, Cybernetics and Informatics  
Volume XIV Computer Science, Engineering Applications  
July 23-30, 2003  
Orlando, Florida – USA  
pp. 391-395

- [RAD99] Radi Amr and Ricardo Poli  
Discovery of Backpropagation Learning Rules Using Genetic Programming  
Genetic Programming, Proceedings of EuroGP'99  
Volume 1598  
May 1999  
pp. 120-134
- [RAM99] Ramesh V. E., M. and Murthy Narasimha  
Off-line signature verification using genetically optimized weighted features.  
Pattern Recognition  
Volume 32,  
Número 2  
Febrero 1999.
- [ROD99] Rodríguez Vázquez Katya  
Multiobjective Evolutionary Algorithms In Non-Linear System Identification.  
A thesis submitted for the candidature of the degree of Doctor in Philosophy.  
Department of Automatic Control & Systems Engineering.  
The University of Sheffield.  
Sheffield, England.  
Año 1999.
- [ROS03] Rosendo R. J. Alejandro, Herrera B. A. Arturo, Padrón G. Alejandro, Prieto M. Rafael  
Algoritmo Genético Multipoblacional y Multicriterio para el Desarrollo de Clasificadores Neuronales Artificiales.  
Memorias del Primer Congreso Mexicano de Computación Evolutiva.  
Centro de Investigaciones en Matemáticas.  
Guanajuato, México 2003  
pp 245-257
- [SM193] Smith Murray .  
Neural Networks for Statical Modeling.  
Primera edición.  
Van Nostrand Reinhold.  
England.  
Año 1993.
- [SUO97] Suenio E, Lipponen P, Mäenpää J, Syrjänen K, Kangas L, Tuomisto L.  
Mitotic index in the subrenal capsule assay as an indicator of the chemosensitivity of ovarian cancer.  
Cancer Chemother Pharmacol  
Springer  
Vol. 41  
1997  
pp 15-21.

TESIS CON  
FALLA DE ORIGEN

TESIS CON  
FALLA DE ORIGEN

REFERENCIAS

- [WEL94] Welstead T. Stephen  
Neural Networks and Fuzzy Logic Applications in C/C++  
Primera edición  
Wiley  
USA  
1994
- [WON98] Wonnacott H. Thomas, Wonnacott J. Ronald  
Introducción a la Estadística  
Quinta edición.  
Limusa-Noriega Editores  
México D.F.  
1998
- [YAO97] Yao Xin and Liu Yong  
A new Evolutionary System for Evolving Artificial Neural Networks  
IEEE Transactions on Neural Networks  
Volume 8  
No 3  
May 1997  
pp 694-713.
- [YAO99] Yao Xin  
Evolving Artificial Neural Networks.  
Proceedings of the IEEE.  
Volume 87.  
No. 9.  
Año 1999  
pp. 1423-1447.
- [ZIT99] Zitzler Eckart and Thiele Lothar .  
Multiobjective Evolutionary Algorithms A Comparative Case Study and the  
Strength Pareto Approach.  
IEEE Transaction on Evolutionary Computation.  
Volume 3.  
No. 4.  
Año 1999.  
pp 257-271.